Data Science Community Newsletter features journalism, research papers and tools/software for August 12, 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.

ADSA ANNUAL MEETING: COVID CONTINGENCY ANNOUNCEMENT

There is a new COVID-contingency plan for the ADSA Annual Meeting. The go/no-go decision regarding the in-person format will be made on or before 1 October. Any registration fees paid before 1 October are eligible for a full or partial refund should the conference move to a virtual format.

THE 6TH IPCC REPORT IS OUT, AND IT'S NOT GOOD NEWS

The IPCC Sixth Assessment Report (AR6) is one big reality check. Evidence from a wide range of sources — paleoclimatology, remote sensing observations, global and regional climate simulations, socioeconomic data, scenarios and observed impacts, emissions data and quantitative scenarios to name a few — published in a vast trove of over 14,000 studies — resulted in an enormous cache of data, which cumulatively are the driving force and evidence behind the report's findings. AR6 is an enormous — and enormously urgent — exercise in data analysis and management. Click through for **more** *From the Desk of Catherine Cramer*.

FACEBOOK SHUTS DOWN NYU RESEARCH INTO POLITICAL ADS, DISINFORMATION

Facebook **shut down** an **NYU** research project that was investigating political ads. NYU researcher **Laura Edelson** points out that even the **Federal Trade Commission** **doesn't think** its consent decree against Facebook requires Facebook to shut the project down. Acting Director of the **Bureau of Consumer Protection Sam Levine** weighed in on the case saying, "I am disappointed by how your company has conducted itself in this matter." Edelson further noted to the DSCN that the entire thrust of the claims Facebook used to shut the project down — that the **Ad Observatory** used improper consenting procedures and/or violated users' privacy — are wholly without merit. The consenting procedures were not only constructed specifically to inform and empower users, they were also **vetted by Mozilla's privacy team**. There's a letter collecting signatures of support **here**, if you want to join the chorus.

GPT-3

There are now two natural language-generating models available to the public(ish): **OpenAI** released **theirs** (the waitlist is still in place) and so did Israeli start-up **AI21** (publicly available, no waitlist).

COVID AND BACK TO SCHOOL, HIGHER EDUCATION VERSION

Proof of vaccination is super antiquated when it relies on a paper card too big to fit in a wallet, too fragile to survive in a wallet for years, and too basic to avoid forgery on the black market. There is a **SMARTHealthIT open digital vaccination standard** for e-vaxx verification that is being used by **Appian** (who will make your company its own vaxx status app), **Google**, **EPIC**, **Cedars-Sinai Health System**, **Mayo Clinic**, **Microsoft**, the **State of California** and now being **adopted** by **New York**. The SMART certificates are technology **originally developed** at **Boston Children's Hospital** (though **Massachusetts** has not adopted the system). It baffles me why more states have not adopted this standard and moved forward with digital vaxx status apps.

Colleges and universities are surprisingly absent from that list given how many are requiring vaccinations. Then again, many are still **struggling about whether they should mandate vaccinations or not**. The **University of California** and **California State** systems along with the **New York State University** system will mandate vaccines. Many, many private colleges are mandating vaccines. But sizable public university systems from Florida to Wisconsin to Texas have not and likely will not mandate vaccines. **Indiana University** announced they would require vaccines earlier this summer, was sued for violating 8 students' Constitutional rights, but has emerged victorious. The **7th Circuit Court of Appeals** **ruled in favor of IU**, noting that students who did not want to get vaccinated could choose to be educated elsewhere. Further, the university has a duty to protect the student body as well as the leeway to determine *how* to provide a safe environment for students. The ruling may not clear a path for public universities that have avoided mandating vaccines to now do so because there are potentially other legal grounds upon which students, staff or faculty could sue though none of those grounds may be strong enough to prevail against the duty of care universities must uphold.

THERE SHOULD BE SOMETHING LIKE UL FOR ML

The **U.S. Department of Homeland Security** released an **18 page report on AI/ML use** that mostly encourages the department to use AI/ML more often and in more ways with an interdisciplinary, cross-cutting approach. They also inflect the document with "responsibility" language and explicitly note that "DHS must develop and field this new technology in a way that protects privacy, civil rights and civil liberties, and protects against bias, both to ensure effectiveness and to maintain public trust." While I'm wary of calls to "do more of way, way more AI (but don't do harm along the way)" because they often result in doing way, way more AI with collateral damage (and worse) along the way, at least the calls to do more AI are now simultaneously accompanied by

an awareness that harms are likely to result in the absence of specific, dedicated, mitigation processes. In the recent past, AI/ML models were used to help predict where people who were in the U.S. illegally could be detained for deportation. These people often had committed no crimes while in the US, were gainfully employed, and were supporting family members (some were even detained at school drop-off/pick-up for their kids). Skepticism about protecting "privacy, civil rights and civil liberties" is warranted here.

On the academic side, there is a new paper in *Nature Machine Intelligence* calling for **third party auditing of all AI/ML models** — the independent audit of AI systems (IAAIS, pronounced 'eyes'). Auditing AI/ML in partially- or fully-automated systems should and could become as normal as UL testing is for appliances and machines that run on electricity. Dear DHS, please put your behavior where your intentions are and submit your AI/ML to third party audits.

TEACHING CS

In August or September we usually provide an annual reminder that there are more undergraduates who want to be computer science majors than there are resources at universities to efficiently support all those CS majors. That ivory tower CS-lane traffic jam **has not cleared**; **New Mexico** has moved the starting line back by requiring high school students to **take at least one computer science course**. That should relieve some pressure on higher ed CS instruction, but here's another piece of good news.

Centered at **Stanford**, led by **Mike Wu**, **Chris Piech**, and **Chelsea Finn** is an extremely data-science-y, open science-y **project** that can educate 12,000 students at a go in a how-to-code class (**Code in Place**). Not only is the core team able to attract an army of volunteer instructors — 1,120 — they have also developed an AI that offers constructive feedback to students as they go through exams. This feedback is not the generic tool-tip stuff or the inscrutable-to-beginners compiler errors. It's feedback designed specifically for the pedagogical environment of first-time coders who are so raw in their enthusiasm for coding, they're able to make an almost infinite number of good guesses. Thinking differently, indeed! But what do students think? After collectively receiving 16,000 pieces of AI-generated advice, "students agreed with the AI feedback 97.9% of the time, compared to 96.7% agreement to feedback provided by human instructors." Pretty cool.

Metascience 2021 will explore the themes of metascience and the scientific process through a global, interdisciplinary, cross-sector lens.

We look forward to engaging with the community of transdisciplinary researchers and stakeholders investigating and shaping the future of science together.

[Tickets are $5 for students/postdocs and $10 for others. Fee waivers are available for those who cannot pay.](#)

SCIENCE!

Quantum computing has been a slow buzz topic, at least in the DSCN. Other outlets have used a few more breathless exclamation points. An astounding research breakthrough involving physics and the chemistry of crystals has discovered essentially a fourth phase of matter — neither solid nor gas nor liquid — that's only possible in a certain type of crystalline structure. In this phase, time becomes less linear (non-linear?) and matter is able to be perpetually excitatory without requiring he energy loss/gain typically associated with a phase change. In other words, this "time crystal" can produce change without increasing entropy, thereby appearing to violate the 2nd law of thermodynamics. Recall that most workarounds for maintaining order despite the 2nd law of thermodynamics usually require expenditures of energy. Where the 2nd law appears not to apply (or not to apply in the way we are accustomed to it applying), it is easy to imagine that there could be computational systems requiring far, far less energy inputs. There's essentially no way I can fully explain both how this works and why it matters in a short blurb — read **the article** by **Natalie Wolchover** in *Quanta* or **the preprint**.

In a completely different field on an unrelated topic though sharing a similar upshot — orders of magnitude less time to insight — a new enzyme kinetic tool speeds up iterative enzyme mutation investigations by ten to one hundred times. "A lot of the industrial chemicals we use now are bad for the environment and are not sustainable. But enzymes work most effectively in the most environmentally benign substance we have — water," said **study** co-first author **Daniel Mokhtari**, a **Stanford** graduate student in the **Herschlag** and **Fordyce** labs, which is a big reason to get excited about enzymes in a general sense. In terms of the specific problems scientists want to address with enzymes now is the fact that we've created tons of plastic that cannot currently biodegrade within any reasonable lifetime. The race is on to discover an enzyme or enzymes that can break plastics into decomposable components; the sooner scientists find those enzymes, the better for everyone.

GOVERNMENTS DO THINGS, TOO

California's attorney general **Rob Bonta** has **begun enforcement** of the most comprehensive single piece of data privacy regulation in the U.S. — CPRA (aka CCPA 2.0). In the crosshairs are cookies that track users from site to site as they look at the internet from with a browser. Companies must offer consumers a "Do not sell my personal information" link that allows consumers to easily opt out of the sale of their data. The definition of "sale" of data can include data that passes from one entity to another without what we may normally consider a typical sale or exchange of money for data. For instance, a retailer may place cookies in a consumer's browser that are then shared with a third party data analytics company. If the retailer fails to disclose the use of cookies for this purpose or fails to provide a choice to opt-out of cookies that are not strictly necessary for the site to function or if the company allows the passage of cookie data for independent use by the third party, they may have received a "cure" letter from CA State Attorney General requiring that they explain their processes and get them into compliance within 30 days.

South Africa and **Australia** will allow AI to be **recognized as co-inventors** in patent applications. The decisions did not address the question of whether AI can have property rights.

Careers

See the for more opportunities.

TIES Assistant Professor or Associate Professor without Tenure

The Massachusetts Institute of Technology (MIT) Sloan School of Management is seeking faculty candidates with research and teaching interests focused on the field of Entrepreneurship, and relevant to the fields of Innovation and/or Strategy for positions in the Technological Innovation, Entrepreneurship and Strategic Management (TIES) Group.

Location: 100 Main Street, Cambridge, Massachusetts, 02142

Applications will be reviewed on a rolling basis as of October 15, 2021. The application deadline is November 12, 2021.

Executive Director, NSF West Big Data Innovation Hub at University of California Berkeley

Location: Berkeley, CA

WEARABLES: OLYMPICS, CONSUMERS AND WHAT'S IN-BETWEEN

The **2020 Olympic Games** of 2021 just concluded. Canadian biomechanics researcher **John Barden** pointed to wearable technology as **a game-changer** for this generation of elite athletes, "They're relatively inexpensive and small" and "provide a vast array of metrics to coaches and athletes in many different sports." Norwegian national coaches **Olav Aleksander Bu** and **Arild Tveiten** turned triathlete **Kristin Blummenfelt** into **a living, breathing laboratory**, something that involved **core body temperature monitoring** among an arsenal of other wearable devices. Blummenfelt **won gold**, sweating completely through his tri-suit in the process, projectile vomiting after the race and leaving the scene in a wheelchair. Affordability makes wearable technology a force multiplier for the best coaches who can help more athletes than before. American universities play a major role as elite Olympic sports democratize globally. The **University of Southern California sent 65 students and alumni** to the Tokyo Games representing the U.S. and 32 other nations. Twenty-one USC athletes earned medals, **more than any other U.S. university**.

Athlete biometrics were also part of the 2020 Olympics' media presentation. Archers **collectively okayed contactless heartrate monitors** for global television coverage. Data ethicist **Timnit Gebru retweeted concerns** that the technology's invasiveness took a backseat to how cool the under-pressure athletes appeared. HIPAA protections **don't apply to athletes' biometrics** though in professional cases, like Olympic archers, workers' rights offer some privacy protection.

Consumer technologies parallel athletes' technologies. Wearables are proving beneficial. **University of Minnesota** researchers **recently showed** that obese individuals who used wrist-worn fitness trackers lost weight and reduced their body-mass index. Contactless sleep monitors from **Amazon** and **Google Nest** recently gained U.S. government approval as health technology, but they are also **controversial**

surveillance technologies. **Google's** purchase of **Fitbit**, one of the first popular activity tracking bracelets, **was finalized** in January, after Google agreed not to incorporate Fitbit data into ads. **Facebook recently announced plans** for its smartwatch with no such privacy commitments or safeguards.

Researchers who want to analyze wearables' data are finding that the consumer technologies are **not reliable measurements**. **Harvard** biostatistician **JP Onnela** wanted to use **Apple Watch** data for a study but found that device algorithms were too inconsistent given the range of feature sets and software upgrades that are offered to customers. **Jacek Urbanek** and **Jennifer Schrack** from **Johns Hopkins University Center on Aging and Health** created **Accelerometry Resource**, the product of **five years spent wrangling wearable device data**. The Resource exists to help peers who want to study wearables but need a hand with study design, device selection, training, results analysis, etcetera.

Something interesting happened when researchers from **Max Planck Institute** in Germany **put accelerometers on a troop of wild baboons** in Kenya's **Mpala Research Centre**. **Rol Harel** and colleagues found that the animals would instinctively adjust their speed to either keep up or slow down with the others. If collective locomotion is a thing among social primates then the African proverb may be neurobiologically true — If you want to go fast, go alone; if you want to go far, go together. **Stanford University** biomechanist **Scott Delp** who directs the new **Wu Tsai Human Performance Alliance** was on **NPR's Science Friday** recently and (unintentionally) **said something similar**. "We're studying [athlete's] peak performance with the goal of enabling all people to achieve optimal health and well-being," Delp told host **Roxanne Khamsi**. Most of what we know about health comes from studying disease, "and we're taking the opposite approach."

FOLLOW THE MONEY
$112,000,000,000 **Infrastructure backlog** facing higher education

$6,500,000,000 **President Biden's Proposed Budget** -> **ARPA-H** (an advanced research project to be managed by the **National Institutes of Health**) [NOTE: this is only a proposed budget line; it hasn't been approved yet.]

$220,000,000 **National Science Foundation** (NSF) -> **11 NSF National Artificial Intelligence Research Institutes** to "to pursue transformational advances in a range of economic sectors, and science and engineering fields"

$41,000,000 Hypersonic Applied Research Facility at **Purdue University** which will enclose **"the only Mach 8 quiet wind tunnel in the world as well as a hypersonic pulse (HYPULSE) shock tunnel."**

$20,000,000 **Westcott Investment Group** -> **Paradromics, Inc.** to **produce a brain-computer interface** designed to give brain-injured people the ability to communicate with words again

$15,100,000 **U.S. Department of Energy** -> 2 National Labs and 5 universities for **3 projects** "to advance the development of a flexible multi-tiered data and computational infrastructure to support a diverse collection of on-demand scientific data processing tasks and computationally intensive simulations"

$10,000,000 **NSF** -> **Duke University** to create **The Alliance for Identity-Inclusive Computing Education** which will provide "access and retention for historically underrepresented groups to computing education"

$10,000,000 **NSF** -> **Auburn University** and 26 other institutions to improve **"STEM education among students with disabilities"**

$3,250,000 **NSF** -> **University of Tennessee, Knoxville** and 10s of other institutions for the **Nuclear Physics of Multi-Messenger Mergers research hub**

$3,000,000 **NSF** -> **University of New Mexico Center for Quantum Information and Control** (CQuIC) to fund **new, prestigous interdisciplinary postdoctoral fellowships** in Quantum Information Science

$2,800,000 **NSF** -> **Rice University** to **"track environmental microbiome dynamics over time, across species"**

$1,5000,000 **DARPA** -> **Charles River Analytics** for NEMO (Nautical Evaluation of Mammal Observation), an autonomous **marine mammal detection system** for crewed and uncrewed watercraft

$410,000 **NSF** -> **Florida State University Department of Scientific Computing** so that **Peter Beerli** and **Somayeh Mashayekh** can **develop new software** that better predicts population sizes and genetic diversity of endangered species

£300,000 **Tencent AI Lab** -> **Oxford University Department of Statistics** for an ongoing large-scale machine learning collaboration despite **increasing scrutiny** on Tencent's role in providing surveillance technology to **the Chinese government**

27-34% return on U.S. college endowments **"best returns in 35 years"**

$0 **U.S. Department of Education** -> **Harvard**, **Yale**, **Amherst**, **Bowdoin**, as these schools **decline the latest round of government aid to universities**, money meant to offset financial impacts from the COVID pandemic

NEW PROGRAMS
Origins of Life Initiative at the **University of Chicago** will seek to understand the earliest processes governing the origin of life on Earth and elsewhere in the universe. Led by Nobel Prize-winner and former **Harvard** professor **Jack Szostak**.

Algorithmic Bias Bounty Program **Twitter** gave $3,500 — 1st place to **Bogdan Kulynych**, **EPFL** (Switzerland) grad student for determining that **younger, thinner women** are favored by Twitter's image cropping algorithm

Unnamed new initiative in public-interest technology at **University of Massachusetts Amherst**, to be led by **Francine Berman** who is joining UMass in the fall

Computer Science + X degree program at **University of Wisconsin-Platteville**, bachelors degrees available for fall 2021 are "CS + Business" and "CS + Supply Chain Management"

Neuro-Computing **The University of Tokyo** and **NTT** to "develop new numerical tools and a simulator for the Coherent Ising Machine"

New School Focused on Climate and Sustainability **Stanford University** "first new school in more than 70 years"

Sports Science PhD Fellowship Program is a collaboration between **Texas A&M Athletics**, **Texas A&M Department of Health and Kinesiology**, the **Huffines Institute for Sports Medicine and Human Performance**, and other on-campus partners.

Data Science Masters at **Loyola University, Maryland** will be fully online.

The National Research Center for Health Disparities a partnership led by **Howard University** and private real estate companies to "attract pharmaceutical companies and biomedical research organizations...focused on finding solutions to chronic illnesses...affecting communities of color"

6G Futures a partnership between **University of Bristol** and **Kings College, London** which will centralize access to "over 400 world-renowned experts in telecommunications networks, cyber, Artificial Intelligence, digital humanities, social sciences and arts to help shape the future of mobile technology for individuals and society"

DATA VISUALIZATION OF THE WEEK
TIME Magazine, Michael E. Mann from August 9, 2021

The latest version of the "hockey stick" chart shows unprecedented warming in recent years.

Change in global surface temperature relative to 1850-1900 average

1.09°C in 2016
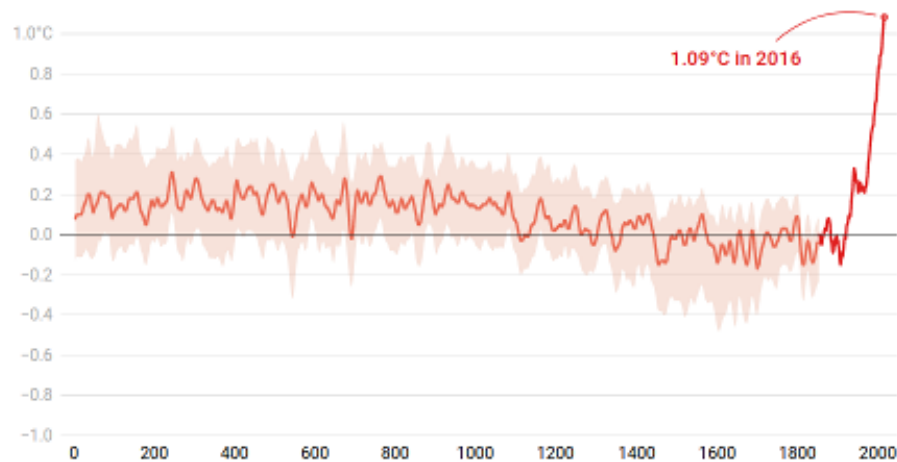
Numbers are observed from 1850–2016; for prior years, they are reconstructed using proxy records like tree rings, corals, and ice cores.

Chart: Elijah Wolfson for TIME • Source: IPCC, 2021: Summary for Policymakers

TIME

Deadlines
Contests/Award
NFL and AWS Launch Artificial Intelligence Challenge to Crowdsource Ways to Automate Player Identification using NFL Game Footage
"New computer vision models created through the challenge will accelerate the **NFL's** work with **AWS** to better understand, and aim to reduce, injuries in the NFL. Prize money totaling $100,000 will be awarded to data scientists with winning models. The contest will be open through November 2, 2021."

Studies/Surveys
Academic researchers, we need your help! Our team [at Twitter] is doing a lit review of published academic papers that have used Twitter data in their analysis, across any field of study.
"Find out more, and see how you can share your work via **Submittable**"

Education Opportunities
HealthTech 2022-23 Fellowship Application Is Open!
"**Harvard HealthTech** is excited to invite eight fellows who will work together on teams of four to join this unique 10-month immersion opportunity in the Boston healthcare ecosystem. Fellows will undergo training in healthcare innovation by immersing themselves in clinical settings to identify unmet healthcare needs, then designing and testing innovative solutions with guidance from mentors in the medtech, healthcare, and innovation sectors." Deadline for applications is September 9.

Tools & Resources

A New Practical Guide to Using Python for Earth Observation
Eos, Rebekah B. Esmaili from August 06, 2021
"*Earth Observation Using Python: A Practical Programming Guide* is a new book recently published in **AGU's** Special Publications series. It presents an introduction to basic Python programming that can be used to create functional and effective visualizations from earth observation satellite data sets."
'More about knowing someone who can get you in the door': Students share advice on landing big tech internships
The Stanford Daily student newspaper, Chuying Huo from August 02, 2021
"Although a prestigious school, solid resume and high GPA can be assets, pure luck can sometimes turn out to be the most important factor in the equation. First-year computer science Ph.D. student **Aaron Mishkin's** success in landing an internship at **Amazon Research** was a matter of being aligned on a research topic with recruiters, getting along well with Amazon faculty members during a research conference and engaging in constructive conversations during interviews."

Events

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