

Please let us ([Laura Noren](#), [Brad Stenger](#)) 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](#) and to NYU's Center for Data Science.

Data Science News

University Data Science News

University of Colorado-Boulder researchers, in conjunction with the **High Plains Research Network**, have been [running a survey on mental health](#) since the pre-COVID days. We all like to think those were better times, which is true!, but they may not have been great times. Only 13.6% of respondents reported "excellent" mental health pre-SARS-CoV-2. That number has now fallen to 6.2%, but if you don't trust anyone who reports excellent mental health, maybe you want to hear about the people reporting "very good" mental health. Pre-COVID it was 36.2% and now it has fallen to 17.8% — a proportionally similar drop. Seventy percent of respondents to the Population Mental Health and Wellbeing Program Survey, which is a different survey on a similar topic, reported being "highly stressed" with one-quarter reporting a "high level" of depressive symptoms and 27% reporting a "high level" of anxiety symptoms. Coloradans in the study are turning towards exercise. Many have found that social networks can be quite helpful, but they may currently be a source of a great deal of friction and fear as those we know get sick, lose their jobs...or choose to party like it's 1999. (I live in Newport Beach the new epicenter of human honey badgers who will stage a showdown with the stay-at-home order this weekend. Mounting anecdotal evidence suggests being outside together is probably not all that dangerous and this weekend will add to that non-consensual natural experiment.)

The **Stanford Daily** student newspaper [visualized trends](#) in undergraduate majors over the past ten years. As expected, the most popular major is computer science, though it was only the third most rapidly growing major. Symbolic Systems (combines math, philosophy, linguistics, cognitive science, neuroscience and a few more subjects sprinkled in) is the fastest growing with psych, chem, and public policy losing students at the fastest rates. I am not sure these are the trends we want to be seeing as we're staring into the face of a national pandemic. But the visualizations are nice.

Jay Bhattacharya, one of the **Stanford** researchers who was a co-author on the **Santa Clara County** and **Los Angeles** antibody tests that received a [heavy criticism](#) for non-representative samples, statistical misunderstandings, and for releasing the LA study to the press rather than as a pre-print is in more trouble this week. It turns out that his wife, **Catherine Su** [sent a recruitment email](#) with the subject line "COVID-19 antibody testing - FREE" that claimed participants would "know if you are immune" and that the test was "FDA approved." The test was not FDA approved and the test does not prove immunity to COVID-19. There is still ongoing debate about whether surviving the

infection offers immunity. These claims are flagrant violations of standard IRB protocols governing how subject recruitment materials can be written. Recruitment letters, emails, and flyers cannot contain outright falsehood (e.g. the test was not FDA approved) and they cannot overstate the benefits to subjects (e.g. the test is not able to prove immunity, as that is a different question altogether). On that point, Su wrote: "If you have antibodies against the virus, you are FREE from the danger of a) getting sick or b) spreading the virus. In China and U.K. they are asking for proof of immunity before returning to work. If you know any small business owners or employees that have been laid off, let them know -- they no longer need to quarantine and can return to work without fear." This is not only an overpromise of benefits to subjects, it also uses fear mongering. For his part, Bhattacharya says he had nothing to do with his wife's email, but admitted that the email did skew the makeup of study participants. A spokesperson for Stanford said only that "these matters are being handled by the appropriate committees."

The test used in the study was not developed by Stanford. It was developed in China, but is not in use there because it has not passed through the Chinese equivalent of the FDA. **Andrew Gelman, Columbia University** professor and statistician [has asked for an apology](#) from the study's authors. We're still waiting.

Across the bay at **University of California-Berkeley** data scientists reviewed data from **Santa Clara County** and **New York City**, finding that the upper and lower bounds on the [true COVID-19 fatality rates for those two locations are probably 0.5% - 1%](#), five times higher than the 0.1% - 0.2% fatality rate claimed in the Stanford-led study on Santa Clara. In **Italy**, the fatality rate is towards the upper end of that range, 0.8%, which is expected due to the larger proportion of older people in Italy. The 0.8% rate is an estimate that could rise if some at-home deaths are reclassified as caused by COVID-19 at a later date. Whether the fatality rate is 0.5%, 0.8% or 1%, it's higher than the fatality rate associated with the seasonal flu (0.1%).

Staying with **Berkeley** news, the university has announced it will [launch](#) a Computational Social Science Training Program for pre-doctoral students with a \$1.2 million grant from the **National Institute of Child Health and Human Development**. Even though it's called a social science training program, it has a strong epidemiology/biostats focus as one of the faculty leads is **Maya Peterson**, an Associate Professor of Epidemiology and Biostatistics.

Four thousand people from the entire San Francisco Bay region will be sampled, tested for COVID-19 and monitored in a collaboration between **Chan Zuckerberg Initiative**, the **Chan Zuckerberg Biohub**, **Stanford University** and **University of California-San Francisco**. The study is part of a [\\$13.6 million effort](#) to understand the spread of COVID-19 across the Bay Area. There will also be a second study to look at protective benefits of COVID-19 antibodies, focusing on healthcare workers.

Rutgers University in New Jersey has also launched [a study of COVID-19 health care workers](#) in the U.S., focusing on those who became ill with the disease. I hope they expand the study to look at the downstream effects on all health care workers, not just those who got sick or even those who treated COVID-19 patients. What kind of trauma scars will they carry? How will their trust in their

hospitals change, if at all? Will some leave the field or seek different specialties?

Tsinghua University in Beijing ranks the [top 2000 researchers in AI](#) by indexing 133 million expert profiles across 270 million publications. This year's list is out and it seems fairly "accurate" to me at first glance. American and Chinese scholars dominate the list with strong showings by Canadian, French, British, German and Japanese scholars. The two people who "won" because they were influential in four sub-fields of AI are **Yoshua Bengio** and **Alex Smola**. Only 9.7% of the influential AI researchers are women. The largest gender gap is in machine learning, which only had 2% of its membership constituted by women. Institutionally, **Google** (185) and **Microsoft** (91) had the largest number of influential AI researchers followed by **Facebook** (59), **Berkeley** (51), and **MIT** (51). It's a comprehensive report with many data visualizations.

The **University of Maryland** has released a [new COVID-19 Impact Analysis Platform](#) which seems more like a dashboard hosting opinionated data about potential contributing factors to COVID-19 caseloads. In this case, opinionated data is good. They appear to be cranking out daily updates on state-by-state behavioral patterns such as trips per person, miles traveled per person, and creating a social distancing index score. It's certainly worth having a look.

This newsletter goes where the news takes us, and these days it's coronavirus pandemic. So too for the **Bill & Melinda Gates Foundation**, which will be [focusing 100% of its attention](#) on COVID-19. Grantees beware: if you want Gates funding, you better figure out how to study COVID-19.

Company Data Science News

Obsidian Security has announced it [will offer security support](#) for the **Zoom** video conferencing platform. The support will help visualize call data to make it easier to spot inappropriate participants — from zoom bombers to corporate spies lurking on boardroom calls — as well as address a host of other potential security and privacy violations.

Reuters is reporting a growing trend among [computer vision startups that are offering reporting services which monitor social distancing](#) at factories, construction worksites and retail locations. This pivot to surveilling employees is unsurprising, but nevertheless alarming. Technologies that march in under the auspices of protecting employees also provide an invasive level of persistent surveillance. It's the panopticon presaged by **Michel Foucault** who wisely noted that we will all become the 'docile bodies' obediently internalizing surveillance culture to serve the needs of states and corporations while sacrificing our own freedoms and more creative expressions of selfhood.

Palantir, a powerful data science company fueled by large contracts with US government agencies including **Immigration and Customs Enforcement (ICE)**, [will operate the data machinery in the federal effort to track coronavirus cases](#). Foucault would have a field day with this one. Palantir was chosen to contract with the **Department of Health and Human Services** on HHS Protect Now. Palantir will be allowed to amass "data from across the federal government, state and local governments, healthcare facilities, and colleges" which is an extremely comprehensive dataset. If U.S. citizens trusted Palantir, this may be a good partnership. But many do not. The company has

been met with protests when it visits campuses to recruit and has been removed from sponsorship of academic conferences, largely due to its participation in the deportation of undocumented residents. Some also object to **Peter Thiel**, it's founder. If you are in the U.S. and do not want your data to make its way to Palantir, you have no choice.

Speaking of location data, **Google** has [released short reports on changes in mobility patterns](#) for most countries in the world and all 50 US states. The .csv's are downloadable, but they have also pumped out pdf's for every area if you just want the headlines. The data were gathered from Android phones where location data was enabled and everything is aggregated so that it would be impossible to pick out information on a single individual. The reports are only available for a short period of time, so if you are interested, go download them now.

Slack has seen [an increase in usage](#) as people work from home. Those users established and then broke several new records for the most simultaneous Slack messages sent in March. What I can add from even more recent data is that March was a crazy month for communication. It took a lot of coordination to move everyone to a WFH posture! From the mid-April data I'm seeing, the uptick in March has been followed by a slow decrease. This is expected. The scholarly communications literature suggests that those who write more email, get more email, creating an accelerating treadmill of communications that they generally come to despise even though they like the people with whom they are communicating. As people see less of each other the literature expect everyone to communicate less. With so much sameness on the day to day level, there's less to communicate about. With fewer incidental conversations in the break room, there are fewer links to funny memes or other follow-ons to send around that keep conversations going. What the literature tells us is that we can expect productivity to increase, which tends to increase satisfaction with one's work, but we may have some work to do to make sure that goals and expectations are shared. [Full disclosure: I have invested in Slack.]

The electronic health record company **Epic** is [using a patient deterioration algorithm on COVID-19 patients](#) to help figure out which ones are most likely to need ICU care. This is the textbook situation in which having an AI assistant ought to be helpful. Doctors' time is limited, ICU space is limited, patients' time is limited, the care regimen varies significantly from patient to patient, and the incidence of COVID cases across the Epic system is far greater than the caseload within any given hospital. However, it hasn't been easy to use the algorithm because patients with COVID have large swings in their deterioration scores from one hour to the next. **Karandeep Singh**, a physician and health informaticist at the **University of Michigan** says "it's the best thing we've got right now to help make decisions" but that "the question of whether this is saving lives is unanswerable right now."

Amazon may emerge from the lockdown with a [dominant retail monopoly status never before seen in American consumer culture](#). Many small retailers and chains will disappear. Amazon happens to be strengthened by the virus. It is also strengthened by gathering data on its early-stage competition who often sell their ware in the Amazon Marketplace. Their killer algorithm learns from, and then burns independent sellers on its platform.

OpenTable shows that [restaurant bookings in the US, UK, Germany, Canada, and Australia fell to zero](#) and have remained there for two months. As most restaurants do not have enough cash on hand to weather a one-month shut down, it is unclear if many will be able to return. The industry reported \$50 billion in losses in April, following \$30 billion in losses for March. Diners may be reluctant to return to restaurants because they cannot wear masks and eat at the same time. Restaurants may die at higher numbers than people due to coronavirus. I'm not suggesting people and restaurants are morally equivalent, but the devastation facing the restaurant industry will alter our public life in the near and medium term in ways that will have deep social and cultural implications.

Government Data Science News

The **European Commission** has **launched** a COVID-19 data sharing initiative — the [European COVID-19 Data Platform](#) — that puts new E.U.-wide data sharing principles into practice. The platform will allow researchers to store and search for data including DNA sequences, protein structures, pre-clinical and clinical trial data, and epidemiological data. The only effort that sort of approaches what the E.U. is doing in the U.S. is the **CORD-19** effort to pull together published coronavirus papers and preprints. It is being run by foundations, corporate partners, and academics, not the U.S. federal government. The strength of a federal approach is likely to outperform the **CORD-19** effort, as it draws on an existing research collaboration structure with all the communications pathways already in place, not to mention the ability to centralize funding outlays. The lack of U.S. federal government leadership on the scientific front has already increased fatalities, but could diminish our ability to return safely, further threatening physical, social, political, and economic health. Presenting an intelligent, well-led, well-resourced response to COVID-19 is critical to the United States' stature as a nation.

An article by **Erin Brown** and **Rebecca Robbins** in **STAT** criticized the [chaotic, competitive approach to fighting COVID](#) that is unfolding in the US which is "lowering all boats" due to lack of coordination and splintered funding streams. If there is going to be a contact tracing app, for instance, there should only be one or the outcome will fail.

Right now, there's an **MIT contact tracing app** in the works, a **U.S. Centers for Disease Control app for health care providers**, the **Apple/Google contact tracing strategy**, and **Mike Bloomberg** has [committed to spend \\$10.5 million](#) on a contact tracing app, mostly for New Yorkers. Also kicking around different parts of the U.S.: **PocketCare+**, a new phone app created by **University at Buffalo**; **Duke University's Covidentify**, an app that connects with heart rate watches; plus ongoing collaborations between researchers and fitness band makers like **Cleveland Clinic** and **Whoop** ([link](#)), and in another case, **Stanford**, **Scripps Research Institute** and **Fitbit** ([link](#)).

After some back and forth, **Germany** will [adopt the decentralized Apple/Google approach](#) with data stored on users' phones, not in a centralized database. **France** and **Britain** are going with the centralized strategy, which gives less power to users to control access to their data. In Britain, where the **NHS** is beloved and (trust in large multi-national tech companies is not nearly as high), [centralizing contact tracing within the NHS makes sense](#).

Checking in on countries around the world, **Portugal** has had widespread voluntary cooperation with the stay-at-home concept even as their manufacturing sector has stayed open. **Portugal's case load (2,403/m) and deaths (95.4/m)** are far lower than neighboring Spain (5,067/m; 519/m). Trust in government and in fellow citizens is also high in **Sweden** where K-12 schools are still in session and there is no mandatory lockdown. **The death rate in Sweden (220/m)** is higher than Portugal, but lower than Spain and Britain. The country is also avoiding the mental anguish, economic burden, and political divisiveness of locking down and fighting over when it is safe to open up again. If you will recall from earlier newsletters, the reason Sweden rejected a lock down is because their chief epidemiologist felt that the COVID-19 fatality rate was way too high in the **Imperial College** model and the **University of Washington, Institute for Health Metrics and Evaluation** model which were dominating other countries' approaches at the time. His model now seems to be more accurate at this point.

What's more troubling is that estimates of deaths due to coronavirus only take into account deaths *known to be caused by coronavirus*, not deaths caused by people who might fail to seek medical care for life threatening issues out of excess fear of contracting COVID at the ER. People with appendicitis have reportedly been waiting until their appendices have burst; people with **symptoms of heart attack** may have chosen to stay home. There are also those whose deaths may have been hastened out of overwhelming mental health conditions that led to overdose or suicide. A new **CDC** study has started to look at **the overall excess deaths in the US** which are up 50% in the 7 hardest hit states, far higher than the currently reported death rate due to COVID-19. Some of the gap is due to people dying of COVID-19 at home or in hospitals where test shortages meant the dead were never tested as well as people dying of some other problem whose lives would have been spared except that they fell ill during COVID-19. It's not straightforward to make these estimates because we cannot simply use the prior year baseline. There are also lives saved due to the unintended positive consequences of COVID-19 such as fewer traffic fatalities and possibly fewer activity accidents with national parks and many state parks closed. Getting to an answer here is extremely important because it will determine how countries choose to open up again. If we are taking an approach that aims to save as many lives as possible with some consideration for the quality of those lives, we owe it to ourselves to understand why people are dying. If people are dying because of the lockdown, not COVID-19, that's a serious consequence indeed. I would also argue that this may be a point to reconsider whether we are collectively comfortable with the number of deaths due to things like traffic fatalities and air pollution. We may have been the proverbial frog waiting around for the water to boil, but now we have a choice. Do we want to send everyone back to offices 5 days a week and on flights to make sales calls and attend conferences if we know those activities will lead to traffic fatalities and shorten life spans due to air pollution? The conversations we're forced to have because we need to decide what we want society to look like could result in major improvements to the status quo if we stage this return intelligently. Some countries will excel, others will fall behind, and we may emerge into a vastly altered hierarchy of nations on the other side.

In **Canada's Ontario province**, police have been given access to a **database of everyone who has tested positive for COVID-19** along with their home addresses and dates of birth. The **Canadian**

Civil Liberties Union has called the move "an extraordinary invasion of privacy."

In **Austria** and **Israel** they are poised to experiment with alternating quarantine (AQ) for school children in which half will go to school Monday-Wednesday, then the other half Thursday and Friday, flipping the following week, starting May 15th. **This model** performed as well as a fully effective 80% quarantine model according to **Baruch Barzel** of **Bar-Elan University**.

Another study from German and British authors led by **Christian Drosten** suggests **kids may be some of the worst asymptomatic spreaders**, though we haven't really been able to test this because schools have been shut down and kids have been well-protected from becoming infected in the first place. Sending them back may endanger some of them, but it could also threaten teachers, parents, and, especially, grandparents (and great-grandparents if they're still alive). That said, because children show fewer symptoms, they don't cough much and have smaller expelled air volume, even though they shed virus at similar rates, they are likely to still be less likely to infect others than an infected asymptomatic or symptomatic adult. We will be watching the Austrian, Swedish, and Israeli case loads closely.

Scientists funded by **World Health Organization** and led by **London School of Hygiene & Tropical Medicine** will be working to keep track of all the different national strategies and **measuring their effectiveness**. Thank goodness. It's complicated!

Funding from **NASA** for several successful, crowd-pleasing Mars exploration missions **has been slashed**. The Curiosity rover stands to see its budget cut by 20%, the Mars Odyssey orbiter would be terminated altogether, and the InSight lander would likely go with it since 60% of the data and 90% of the commands to the InSight come through the Odyssey orbiter. There's no coherent reason given, so there is hope that **Congress** might restore funding in one of its funding bills. That's all Congress does these days — approve funding bills — so there is reason to root for the Mars mission. If we can't have sports, restaurants, dates, church, office snacks, classrooms, museums, concerts, or movie theaters, at least we can preserve access to Mars data!

It is imperative that research on COVID-19 be published quickly and made freely available to scholars around the world in order to successfully mitigate the threat. Unfortunately, **China appears to be implementing a censorship program** referred to as "pre-publication vetting for COVID-19" in which government officials will determine if the 'academic value' and the 'timing is right' for publication. This threatens to jeopardize scientific advances as over one-third of the open access studies that have been published about COVID-19 include Chinese authors. Publishers will be put in a difficult position, according to **Taylor & Francis** employee **George Cooper**. He explains that, "publishers would be confronted with a binary choice: remove 'sensitive' Open Access articles from their Chinese platforms, or risk huge losses of revenues and access for their entire published output." This story is developing as we learn more about the Chinese restrictions.

The **U.S. Patent and Technology Office** (USPTO) has ruled that AI cannot invent anything; **only humans can apply for patents**. This only makes sense. The USPTO has seen what bots have done

to **Twitter** and decided they could not work under such conditions. OK, I may have completely fabricated their reasoning, but the facts remain: no bots allowed as authors on patents.

Extra Extra

That whole weird notion that **5G causes COVID-19** is a misinformation campaign propagated largely by bots, according to research by **Kathleen Carley** of **Carnegie Mellon**. The origin of the bots is still unclear.

		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

FEB

					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	

MAR

						1
2	3	4	5	6	7	8
9	10	11	13	12	14	15
16	17	19	18	20	21	22
23	24	25	26	28	27	30

31	30
----	----

APR

						1	
						2	x
						5	10
						8	11
						17	17
6	7					10	19
13	14	15	16	17	18	20	19
						9	19
						30	

@instachanz

This image was tweeted by CMU affiliate Scott B. Weingart and captures the essence of the cultural

moment.

Data Visualization of the Week

Dr. Alex Hanna / Kate Silver liked

stuart andrew thompson @stuartathompson · 6h
recently became *obsessed* with "experts" who say we could have a vaccine in 12 to 18 months. how is that even possible?

i sketched out all the steps here:
[nytimes.com/interactive/20...](https://www.nytimes.com/interactive/2020-04-30/health/coronavirus-vaccine-timeline.html)

The chart displays a timeline from 2020 to 2040. A vertical red line is positioned at May 2036, labeled 'Vaccine by May 2036'. The stages and their durations are as follows:

Stage	Start Year	End Year
Academic research	2020	2022
Pre-clinical	2022	2024
Phase 1 trials	2024	2025
Phase 2	2025	2026
Phase 3	2026	2028
Building factories	2028	2034
Manufacturing	2034	2036
Approval	2036	2037
Distribution	2037	2038

0:06 | 31.8K views

38 | 255 | 438

Twitter, stuart andrew thompson from April 30, 2020

Tweet of the Week



 To my friends at [@Google](#) working on Google Slides:

- Bullets are your primary use case!
- But they are infuriating
- The bullets don't line up
 - They change shape
 - Tab creates these monsters
 - And these
- Please
 - Just make it work



9:33 PM · Apr 28, 2020 · [Twitter Web App](#)

207 Retweets 3K Likes

Twitter, Matt MacInnis from April 29, 2020

EVENTS

[1/ Spent the last couple weeks in quarantine obsessively coding a website for Virtual ICLR with @hen_str . We wanted to build something that was fun to browse, async first, and feels alive.](#)

Online April 26-30. "Here's a sneak peek at a part of the conference portal: switch your watching next week and scroll for papers, videos and discussion instead." [\$\$\$]

[ICLR | 2020 - Eighth International Conference on Learning Representations](#)

Online April 26-30. "We have released a call for virtual rooms and socials that makes it possible for everyone to more easily meet." [\$\$\$]

[Women in Data Science Puget Sound 2020 Conference](#)

Online May 4, starting at 9:30 a.m. PDT. [free, registration required]

[#HashtagActivism](#)

Online May 5, starting at 12:30 p.m. EDT. "Join authors **Sarah J. Jackson**, **Moya Bailey**, and **Brooke Foucault Welles** to look at how marginalized groups use **Twitter** to advance counter-narratives, preempt political spin, and build diverse networks of dissent." [free, registration required]

[Online Causal Inference Seminar](#)

Online May 5, starting at 8:30 a.m. PDT. Speaker: **Eric Tchetgen (Wharton)**. "To stay up-to-date about upcoming presentations and receive Zoom invitations please join our mailing list."

Physics in ML virtual talks

Online May 6. Natural Graph Networks. Speaker: **Taco Cohen, Qualcomm AI Research**. "For link and password to the talks, please sign up for the Physics \cap ML mailing list."

AIIDE 2020 - The 16th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

Online October 19-23. " Due to COVID-19, the organizing committee has decided that AIIDE'20 will be held online. The conference dates remain the same, as shown below. As usual, accepted papers will be published by AAAI and at least one of the authors must register and attend the conference to present their work." Deadline for paper abstracts is May 29.

DEADLINES

Contests/Award

Sounders FC Data Analytics Competition

"**Sounders FC** is challenging you to use data visualization to tell a story and deliver insights that help decision-making process of key stakeholders at club." Deadline for submissions is May 10.

Space Security Challenge 2020 HACK-A-SAT

"The **United States Air Force**, in conjunction with the **Defense Digital Service**, presents this year's Space Security Challenge, Hack-A-Sat. This challenge asks hackers from around the world to focus their skills and creativity on solving cybersecurity challenges on space systems." Online event takes place May 22-24.

KDD Cup 2020

"The KDD Cup competition is anticipated to last for 2-4 months, and the winners will be notified by mid-July 2020. The winners will be honored at the KDD conference opening ceremony and will present their solutions at the KDD Cup workshop during the conference."

Conferences

KDD Humanitarian Mapping Workshop

San Diego, CA, and Online August 24. The workshop "brings together a global community of leading researchers and decision-makers from computer scientists, data scientists to epidemiologists, economists, urban policy researchers, computational social scientists, operations researchers, privacy researchers, legal scholars, and humanitarian organizations to advance the field of crisis informatics with a commonly shared priority research agenda." Deadline for participation is May 20.

SocInfo 2020

Pisa, Italy, and Online October 6-9. "Considering the current circumstances, the impact of the pandemic and ongoing travel bans, organizers of SocInfo2020 are moving forward with the

assumption that SocInfo2020 will be a hybrid of in-person and online conference or a fully online conference probably in the originally scheduled dates of October 6-9, 2020." Deadline for abstracts is May 29.

Education Opportunities

Data Science for Science Teachers Boot Camp

"The **National Institutes of Health's Office of Data Science Strategy** is offering a Data Science Boot Camp for science teachers interested in learning how to incorporate data science into the classroom. The camp will be held July 6-10 in Shady Grove, Md." Deadline to apply is May 1.

A New Fellowship for Tech + Society Strategists

"**Mozilla** and [**Ford Foundation's**] 'Tech & Society Fellowship' is a 24-month program supporting five to ten tech-and-society strategists across the Global South. Fellows will embed with existing civil society organizations to recognize, design, and implement a strategy that brings together a specific regional issue and technology. Fellows will be developers, tech policy analysts, designers, and other tech-specific professions. And they will build tools, grow communities, and conduct research." Deadline for applications is June 2.

Social Data Research and Dissertation Fellowships

"The Social Data Research Fellowship and the Social Data Dissertation Fellowship, new endeavors of the **Social Science Research Council**, with support from **Omidyar Network**, seek to encourage multifaceted pathways for the collection and analysis of social data, with the larger aim of cultivating robust research on technology and society. In particular, we are interested in supporting research that makes creative use of available social data to investigate how social media interact with democracy and elections." Deadline for applications is June 16.

Studies/Surveys

VPRO AI Song Contest

"In the AI Song Contest, teams of musicians, artists, scientists and programmers explore the creative power of artificial intelligence (AI) and the influence AI will have on the music industry in the future. A total of 13 teams from Australia, Sweden, Belgium, The UK, France, Germany, Switzerland and The Netherlands are competing." Voting online until winner is announced on May 12.

RFP

Google News Initiative Innovation Challenges - North America

"The Challenge will focus on projects that generate growth and diversification of revenue for local media who elevate underrepresented audiences and promote diversity, equity and inclusion (DEI) within their journalism. **Google** will fund selected projects up to USD \$300,000 and will finance up to 70% of the total project cost. Special discretion on the total project cap may be considered by The Jury depending on the scale and impact of a very large collaborative application." Deadline to apply is August 12.

Tools & Resources

Colorizing a Visualization

Medium, Nightingale, Theresa-Marie Rhyne from April 02, 2020

"Using Harmony and the Albers App to achieve color harmony in your data visualizations"

A Taste of WebGPU in Firefox

Mozilla Hacks, Dzmitry Malyshau from April 23, 2020

"WebGPU is an emerging API that provides access to the graphics and computing capabilities of hardware on the web. It's designed from the ground up within the W3C GPU for the Web group by all major browser vendors, as well as Intel and a few others."

Amazon AppFlow

Amazon Web Services from April 22, 2020

"a fully managed integration service that enables you to securely transfer data between Software-as-a-Service (SaaS) applications like **Salesforce**, **Marketo**, **Slack**, and **ServiceNow**, and AWS services like **Amazon S3** and **Amazon Redshift**, in just a few clicks."

SPECTER: Document-level Representation Learning using Citation-informed Transformers

arXiv, Computer Science > Computation and Language; Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, Daniel S. Weld from April 20, 2020

"We propose SPECTER, a new method to generate document-level embedding of scientific documents based on pretraining a Transformer language model on a powerful signal of document-level relatedness: the citation graph. Unlike existing pretrained language models, SPECTER can be easily applied to downstream applications without task-specific fine-tuning. Additionally, to encourage further research on document-level models, we introduce SciDocs, a new evaluation benchmark consisting of seven document-level tasks ranging from citation prediction, to document classification and recommendation. We show that SPECTER outperforms a variety of competitive baselines on the benchmark."

Facebook AI, AWS partner to release new PyTorch libraries

Facebook Artificial Intelligence from April 21, 2020

"As part of the broader PyTorch community, **Facebook AI** and **AWS** engineers have partnered to develop new libraries targeted at large-scale elastic and fault-tolerant model training and high-performance PyTorch model deployment. These libraries enable the community to efficiently productionize AI models at scale and push the state of the art on model exploration as model architectures continue to increase in size and complexity. Today, we are sharing new details on these features."

Making Decision Trees Accurate Again: Explaining What Explainable AI Did Not

The Berkeley Artificial Intelligence Research Blog, Alvin Wan from April 23, 2020

"The interpretability of neural networks is becoming increasingly necessary, as deep learning is being adopted in settings where accurate and justifiable predictions are required. These applications range

from finance to medical imaging. However, deep neural networks are notorious for a lack of justification. Explainable AI (XAI) attempts to bridge this divide between accuracy and interpretability, but as we explain below, XAI justifies decisions without interpreting the model directly."

So, you want to run an online experiment?

SAGE Ocean, Jim Lumsden from February 19, 2020

"This post will explore some of the tools and platforms that can help with a key stage of the online research process: creating your survey or experiment. Specifically, we'll be looking at options for running online experiments, with a slight focus on the more complex platforms – those designed to collect reaction time data (e.g., cognitive tasks), or to deliver complex experimental paradigms with a range of response types."

NEW COVID-19 Data Hub V 1.0

U.S. Census Bureau from April 20, 2020

"This site provides users demographic risk factor variables along with economic data on 20 key industries impacted by Coronavirus. Each data set can be displayed in different visualizations, maps, can be shared, and available for download."

Resources for COVID-19 | China Data Lab

Harvard University, Institute for Quantitative Social Science from April 01, 2020

"This project aims to provide an information infrastructure for the spatial study of the new novel coronavirus (COVID-19), which was first reported in **Wuhan, China** and then were found in more than 60 countries and regions in the world."

CAREERS

Full-time, non-tenured academic positions

Senior Lecturer in Data Science

University of Greenwich, School of Computing & Mathematical Sciences; Greenwich, England

Director of Research Data Governance (reports to associate vice provost for research policy and integrity)

Stanford University, Office of the Vice Provost and Dean of Research; Palo Alto, CA

Assistant Dean for Marketing and Communications

University of Illinois at Urbana-Champaign, Grainger College of Engineering; Champaign, IL

Associate Dean for Academic Engagement

James Madison University, JMU Libraries; Harrisonburg, VA

Associate Dean for Discovery, Access, and Technology

James Madison University, JMU Libraries; Harrisonburg, VA

Postdocs

Postdoctoral Fellow

University of Hong Kong, School of Biomedical Sciences; Pok Fu Lam, Hong Kong

Research Fellow, Biostatistics

Dana-Farber Cancer Institute, Harvard Medical School; Boston, MA

Postdoctoral Researcher, Machine Learning

Pfizer, R&D; Cambridge, MA

Postdoctoral Scholar Openings

Stanford University, Artificial Intelligence Laboratory; Palo Alto, CA

Postdoctoral Fellowship in Computer Science: Knowledge Graphs and Natural Language Processing

Stanford University, Stanford Network Analysis Project; Palo Alto, CA

Covid-19 Behind Bars Data Project Research Fellows (2)

University of California-Los Angeles, Prison Law and Policy Program; Los Angeles, CA, or Remote

Postdoc in Computational Social Science

Northeastern University, Network Science Institute; Boston, MA

Full-time positions outside academia

Senior Data Visualization Analyst

Pareto Intelligence; Chicago, IL

Senior Epidemiologist/Study Director

Westat; Atlanta, GA

Data specialist (contract)

The Washington Post, Advanced Storytelling Lab

Data Analysis Team Manager

Space Telescope Science Institute; Baltimore, MD

Internships and other temporary positions

Text Mining Student Specialist

University of Washington, eScience Institute; Seattle, WA

NIH DATA Scholar - Data Harmonization, Mobile Analytics, and End-User Support

National Institutes of Health, Office of Strategic Coordination; Bethesda, MD

Teaching Specialist – Data Science and/or Data Management

University of Minnesota, College of Continuing and Professional Studies; Minneapolis, MN

Technical Writer

Code for Science & Society Inc, OpenRefine; Remote

Scientific Software Developer- Contract Basis

NumFOCUS, SunPy Project; Austin, TX

Click here to receive the Data Science Community Newsletter and/or to have us follow your twitter feed so that our data science twitter bot can easily grab links from your tweets.

To send us an announcement for the newsletter, please email laura.noren@nyu.edu and brad.stenger@gmail.com.

We retain curatorial discretion.

Data Science Community Newsletter Issue 200.