



Data Science Community Newsletter features journalism, research papers and tools/software for March 31, 2021.

Please let us ([Micaela Parker](#), [Steve Van Tuyl](#), [Catherine Cramer](#), [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](#).

From the desk of Laura Norén

The field of cybersecurity – which has given up its training wheels and is simply called "security" now – has seen an increase in the number of vendors claiming to use AI to detect and/or prevent malicious intrusions and attacks. AI in this context is incredibly seductive – wouldn't it be great if an algorithm cranking through real-time and historical audit logs, user activity, and user status information could effectively head off bad situations before they turn into big problems? But there are five reasons why we should resist this seduction which are (mostly) specific to security. More at <https://academicdatascience.org/resources/community-blog>.

Academic Data Science News

The ways we all work have all been impacted by COVID – including students' academic performance. Several online tools such as **Chegg** and **Course Hero** have led to a dramatic increase in [what universities view as cheating](#) – resulting in course-correction by the tool developers, but also reflects students' attitudes toward remote learning – a resentment-tinged "if you're not going to do your job I'm not going to do mine" being [one of them](#).

WFH for the rest of the working world is also a mixed bag, negatively affecting our [workplace relationships](#) and rates of inclusion (particularly affecting women and parents of young children),

leading companies of all sizes to consider [best practices](#) for bringing people back to the office post-COVID. A global survey of 500 large companies by accountants **KPMG** reveals [less than half](#) expect a “return to normality” in 2022 – though (good news for commercial real estate) the expected widespread abandonment of commercial office space does not appear to be happening.

One thing that hasn't changed? The [horribleness of job interviews](#), according to the new book *Noise: A Flaw in Human Judgment* by **Daniel Kahneman**, **Cass Sunstein** and **Olivier Sibony**. The culprit? Unwanted variability of judgments, a.k.a. “noise” causing “human decision-making gone off the rails.”

The “app” is now a cornerstone of students' classwork, and increasingly apps have become a force-multiplier for student learning and expression. The classic team term project app can, by itself, bear fruit, [like the collaboration](#) between **Cornell** students and **Ithaca town government** to shovel sidewalks after snowstorms. Other times an app's starting point is the campus entrepreneurship hub that brings together like-minded talents, like this **University of Toronto** [urban restaurant matching app](#), designed to help small businesses that have struggled with the pandemic. University-generated apps can scale to global levels. Founders of the startup **Discover Together** [found each other online](#) as students at **University of Southern California**, **UCLA**, **University of Waterloo** and **Cornell** did, and now have a promising app that transforms music finding on Spotify from a singular pursuit to something that couples can do together. Associate professor **Robb Rutledge's** neuropsychology lab at **Yale** has perfected [the app as experimental platform](#), and has recently launched “The Happiness Project,” a tool for studying decision-making and well-being.

New York Times AI beat reporter **Cade Metz** has a new, humbly-titled [book](#), *Genius Makers: The Mavericks Who Brought AI to Google, Facebook, and the World*. It profiles the original creators of neural nets: **Yann LeCun**, **Geoff Hinton**, and **Yoshua Bengio**. **WIRED** has [an excerpt](#) mostly focusing on the time when Geoff Hinton turned himself and two of his students into a company in order to auction themselves off to the highest bidder. And if you're a LeCun fan, Metz will be doing [a fireside-Zoom](#) with Yann presented by **NYU** on April 9th.

Public health schools are seeing [a flood of new applicants](#). **STAT**, the medical news service, calls it “the **Fauci** effect.” The increases are significantly greater among Black and Latino candidates, evidence that people see a link between social justice and equal healthcare opportunity. Still other near-term trends discourage. This year's MD fellowship match saw

dozens of infectious disease openings go unfilled, meaning infectious disease specialists will continue to be lacking in numbers.

The **COVID-19 Prevention Network**, based at **Fred Hutchinson Cancer Research Center**, is [currently enlisting 12,000 college students](#) at 21 U.S. campuses (**University of Colorado** is [one](#)) to investigate if the **Moderna** vaccine prevents spread of the infection. The study is funded by the **National Institute of Allergy and Infectious Diseases**. Early evidence, including [a new study](#) by the **U.S. Centers for Disease Control**, shows that vaccination does reduce infection, including asymptomatic infection. The study also shows the increasing supply of vaccine that is available for people under age 65, enough supply that universities have begun to follow through on [plans for large-scale inoculation centers](#). But state-managed websites are proving to be a bottleneck, sometimes with [loooooong page load times](#). In **New Hampshire**, the governor is pleading for patience. **Rutgers** will [require vaccine proof](#) for students to attend classes on campus in the Fall. University researchers have published [guidelines](#) for managing vaccine prioritization and distribution through the **Computing Research Association**.

North Carolina State University just launched its [university-wide Data Science Academy](#). No acronym, just Data Science Academy. It will be the school's hub for cross-departmental coursework, for data science consulting, for tutorials and workshops and for long-term collaborative research engagements. Eventually there will also be an interdepartmental undergraduate data science minor.

University of California-Davis has a [new undergraduate data science major](#) that will start in the Fall of 2022. **University of Massachusetts Amherst** [announced](#) its new online Master of Science Program in Computer Science.

The **Dallas Morning News** [long-formed Southern Methodist University's](#) investments in data science research and infrastructure with a sponsored post to the paper's FWD DFW blog.

Syracuse University School of Information Studies [updated curriculum requirements](#) for the Applied Data Science master's degree. The program will emphasize portfolio work and will have three specialization tracks: Data and Business Analytics, Language Analytics, and Data Pipelines and Platforms.

Lots of partnering going on. The **Boston University Center for Antiracist Research** and the **Boston Globe's** Opinion editorial team [will co-produce](#) *The Emancipator*, a publication

modeled on 19th Century abolitionist newspapers like **William Lloyd Garrison's** *The Liberator*. **Carnegie Mellon** and the **University of Pittsburgh** are [launching](#) a joint Collaboratory Against Hate - Research and Action Center. **Ford Motor Co.** and the **University of Michigan** are sharing [a new 134,000-square-foot robotics research building](#) in Ann Arbor. The bottom three floors are UM and Ford gets the penthouse. The **University of Alberta** and **TELUS** have a new \$15 million ["living laboratory"](#) to research 5G infrastructure and develop applications for precision agriculture and autonomous vehicles. **Johnson Controls**, a multi-national building controls conglomerate with headquarters in Ireland (founded in Wisconsin) is collaborating with **National University of Singapore School of Design and Environment** on a \$5 million [effort to advance smart buildings research](#). The **Halicioglu Data Science Institute** at the **University of California-San Diego** has a small but growing [data sharing partnership](#) with **Intel** that recently expanded to include six HDSI faculty and nine students across seven different projects.

Switzerland's **EPFL** approved creation of a [new Center for Imaging](#) on campus that will further the school's original Imaging@EPFL initiative that started in 2019.

Simon Sheather [has been named](#) the 1st Endowed Truist Chair in Data Analytics at the **University of Kentucky**. It's also a first for DSCN, reporting data science news from UKY and Lexington. Sorry we've overlooked you for so long.

SPONSORED CONTENT



The **UVA School of Data Science** aims to further discovery, share knowledge, & make a positive impact on society through collaborative, open, and responsible data science research & education. We offer residential & online MSDS [degree programs](#) as well as [executive education programs](#). We want to educate students in data science throughout the academic pipeline & are currently developing more academic programs. To stay in the loop & to hear about hiring opportunities, subscribe to our [monthly newsletter](#).

Editor's Picks

Our collective COVID-19 status is precarious. To grasp where we really are, looking at [hospitalizations](#), [viral reproduction number and vaccination rates](#) offers more accuracy. Much of what has gone wrong traces back to ill-informed processes leading to inaccurate data – woeful [government technology](#), inadequate [data tracking](#), erroneous [Excel spreadsheets](#), good old [SARS-CoV-2 case miscounts](#) (and [more miscounts](#)). Current outbreak-causing behavior dilemmas include [youth sports](#) and [increased spring break travel](#). Not very home-stretchy. But being human we do move forward, learning lessons on how to measure population prevalence in epidemics, gaining ground on [telemedicine](#), [long-haul treatment and better testing](#) and figuring out the [legal and privacy implications](#) of (potentially) valuable uses of AI and ML tools in predicting and preventing future pandemics.

In a promising move for those of you who want to use AI and data science for good, the [Patrick McGovern Foundation](#) and the [Cloudera Foundation](#) announced that they [have merged](#). Both foundations had established their missions as 'tech for good', but it is still quite unusual for non-profits to merge. Together, the two will have an endowment worth \$1.5 billion, plus meaningful experience in data science, and they have committed to using data science to achieve the UN's sustainable development goals. Somewhat awkwardly, the McGovern Foundation's main program area is called Data and Society, which is the same name as the [New York-based think tank](#) established by **danah boyd** in 2014.

The past, present and future of the Internet are intertwined, especially for researchers. And right now the future of the Open Internet is uncertain, subject to nation-states' government firewalls, corporate walled gardens and an in flux U.S. regulatory environment. The **Internet Archive**, host of the Wayback Machine and longtime Internet librarian, [continues to fight balkanization of the Web](#). The **Free Software Foundation** saw the departure of [basically the entire management team](#) in the wake of major tensions around founder **Richard Stallman**, who has been repeatedly accused of sexual harassment, though he was recently allowed to rejoin the FSF board. Even with the opacity of **Facebook's** walled garden, it is possible to gain an understanding of what's going on. **The Markup** [recently published](#) work from its Citizen Browser project that shows what Facebook's recommendation algorithms shows to different categories of users, for example Biden Voters versus Trump Voters.

Today's choices affect tomorrow's Internet. Non-profit data and technology consultancy **The Engine Room** has pledged to prioritize "Technology in the service of justice" and to choose tech that matches its mission. The justice-focused technologies are, according to the organization, "computer-assisted tools, systems, methods and processes that intentionally weave anti-oppression strategies into all stages of their creation, maintenance and uptake." At the student-consumer level, media and government officials are trying to rein in widespread use of the **Sci-Hub** republisher of academic papers.

Bias in AI-supported tools and software remains pervasive. In facial recognition, wearing a white mask might be a temporary fix, but of course doesn't solve the problem. COVID-19 has had disproportionately more severe outcomes in poor and non-white communities, and myriad forms of bias in datasets used to train, test, and validate AI models add to these disparities – when data from underserved communities are excluded, by biased note-taking in clinical settings, and the effects of skin pigmentation on pulse oximeter accuracy – to name a few among many. The vast majority of DNA represented in genetic databases comes from white people – evidence of systemic racism and inequality. As Vera Jourova, a European Commission Vice President, puts it, copy-pasting the biases of the real world onto the AI-driven world needs to be rigorously avoided, and the EU seems poised to make progress.

Now if we could also solve the other big AI problem – outsize energy use (i.e., greenhouse gas emissions) – then we'd be getting somewhere.

Research News

Big data. **Texas A&M** computer science researchers recently presented an algorithm to rank large datasets' features in one pass. The technique uses a cost function to precisely track features' relative importance. Work needs to be done to advance the algorithm from one-dimensional data to more complex structured data types. **NYU Center for Data Science** Faculty Fellow **Angela Radulescu** is a computational psychiatrist who parses big data to individuate causes of mental illness from the cognitive and behavioral symptoms and from a vast range of environmental factors affecting each case. Her approach is to link natural language processing to personal narrative with an eye toward future treatments involving mixed reality and virtual reality environments.

The **National Science Foundation's** Long Term Ecological Research (LTER) program has been around for over 40 years and has a network of over 1800 researchers at 28 sites, but

the [newest site is unique](#): its goal is to study the urban environment of **Minneapolis-St. Paul** and how the region's residents interact with the ecosystem, both natural and manmade. While other sites might be gathering data on the effects of acidification or habitat loss on regional flora and fauna, this LTER will be looking at how turmoil, socioeconomic disparities, pollution, habitat loss and climate change interact, bringing climate change and socioeconomic disparities together to gain understanding of social-ecological responses to environmental and social changes and test whether theories and models developed in non-urban ecosystems can predict patterns in urban systems.

Researchers on a **Schmidt Futures-backed** project will use machine learning to address the imprecision of existing models to [more holistically capture](#) physical ocean, ice, and atmosphere processes. Existing climate models use relatively large-scale cells or grid boxes to mimic how systems interact; processes that happen on scales that are smaller than the grid cell, e.g. clouds, turbulence, and ocean mixing, are not well captured. Machine learning will guide the development of models of these complex processes directly from data for use in global climate simulations, addressing model shortcomings affecting accurate forecasts of weather and climate on regional and global scales.

Tweet of the Week

Twitter, Kareem Carr from Mar 27, 2021



Data Visualization of the Week

Twitter, Geoff Mulgan from Mar 18, 2021



Events

See the [ADSA Events Page](#) for more details and more opportunities.

Wrangle Summit 2021 Online April 7-9, organized by **Trifacta** and **Google Cloud**. "The first industry event focused on data engineering" [registration required]

New York Times' Cade Metz and NYU's Yann LeCun Discuss AI's Visionaries—April 9 Webinar Online April 9, starting at 1 p.m. Eastern. "**Cade Metz**, a **New York Times** technology correspondent, and **NYU's Yann LeCun** will discuss Metz's newly released *Genius Makers: The Mavericks Who Brought AI to Google, Facebook, and the World*." [registration required]

An exciting lineup of workshops this year at MLsys! Online April 9 starting at 5 p.m. Pacific. "Ranging from on-device ML and graph learning systems to open discussions about unexpected limitations, retrospectives, and negative results in MLsys research!". [registration required]