

Data Science Community Newsletter

18 August 2020

Data Science Community Newsletter

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

Academic Data Science News

[National Science Foundation Invests \\$104 Million To Launch Four New Engineering Research Centers](#) (Forbes, Michael T. Nietzel)

[Stanford launches new master's degree program in education data science](#) (Stanford University, Graduate School of Education)

[Stanford Center for Health Education Launches Online Program in Artificial Intelligence in Healthcare to Improve Patient Outcomes](#) (PR Newswire, Stanford Center for Health Education)

[Harvard Data Science Initiative Awarded Grant from the Alfred P. Sloan Foundation](#) (Harvard University, Harvard Data Science Initiative)

[University Creates Institute for Data Science](#) (University of Mississippi, Ole Miss News)

[Data Science Bachelor's Degree First of Its Kind at University of Arkansas](#) (University of Arkansas, News)

[Mellon Foundation grant supports development of a plan for using artificial intelligence to plumb the National Archives](#) (Virginia Institute of Technology, Virginia Tech Daily)

[Research Slated for Fall Will Stumble Without Undergraduates](#) (The Scientist Magazine®, Amanda Heidt)

Editor's Picks

The ground is shifting beneath the feet of universities and their student customers. **Nature** just called [the situation](#) "a vast unplanned pandemic experiment." [Ivy League schools](#) have largely moved away from [on-campus instruction](#), consistent with the [larger trend among U.S. universities towards online-only](#). Still, there are more than a few universities that will [stick to in-person reopening plans](#) despite local objections and peer schools' doubts.

Computer scientists at **UMass Amherst** are offering [a wifi-based contact tracing app to other schools](#) who want a better handle on a campus' human-human interaction amidst this pandemic. The **Alabama University System** turned to the health professionals at **University of Alabama-Birmingham**, the **Alabama Department of Health** and tech contractor **MotionMobs** to develop [GuideSafe](#), a contact tracing app that leverages the **Google-Apple** exposure notification system. GuideSafe is not the first app to debut using the Google-Apple APIs; [that distinction](#) goes to the **State of Virginia**. The nation of **Greece** is collaborating with **University of Southern California** researcher, **Kimon Drakopoulos**, on [community-level contact tracing](#) that requires lower levels of user uptake. But [privacy concerns loom](#), especially among campus communities, where some see the apps as invasive and an inadequate substitute for trained human beings.

Evidence-based journalists are rising to the occasion in this summer of social justice. **Jeffrey Dastin** at **Reuters** looks into [the facial recognition system that Rite Aid drug stores deployed in U.S. stores](#), "In areas where people of color, including Black or Latino residents, made up the largest racial or ethnic group, Reuters found that stores were more than three times as likely to have the technology." **Juliet Isselbacher** at **STAT** finds that [nearly half of low-income communities have no ICU beds](#) in their area. **Anna North** at **Vox** puts a spotlight on [the maternal health care crisis affecting Black moms](#) that is a byproduct of the pandemic. And a team of reporters at **WBEZ** public radio in Chicago spoke to relatives of COVID-19 victims and connected those insights to data in order to [understand the pandemic's disproportionate impact on minorities](#).

In case you missed **Ed Yong's** [expose](#) in **The Atlantic**. Also, please read **Mary Beth Griggs'** weekly update on COVID-19 research at **The Verge**. Last week she discussed the [high-profile Internet back-and-forth about accelerating COVID vaccine trials](#), a back-and-forth that ended with computer scientist **Steven Salzberg** [admitting on Twitter](#) that he was wrong to say that skipping Phase 3 Vaccine Trials should happen.

Research News

[NSF grant changes raise alarm about commitment to basic research](#) (Nature, News, Giuliana Viglione)

[NIH harnesses AI for COVID-19 diagnosis, treatment, and monitoring](#) (National Institutes of Health (NIH), News Releases)

[Senior U.S. lawmaker wants National Academies to scrutinize racism in science](#) (Science, Jeffrey Mervis)

[Department of Energy Announces \\$8.5 Million for FAIR Data to Advance Artificial Intelligence for Science](#) (U.S. Department of Energy)

[NSF Funds Five New XSEDE-Allocated Systems](#) (The Extreme Science and Engineering Discovery Environment (XSEDE))

[Forecasting Costs for Biomedical Data Preservation](#) (Vimeo, The National Academies)

[With a nod to UC Berkeley, Google crowdsources earthquake data](#) (University of California System, UC Berkeley)

[NSF recognizes innovative science of chemist and aeronautical engineer as 2020 Waterman awardees](#) (National Science Foundation, News Release)

[Shrinking deep learning's carbon footprint](#) (MIT News, MIT Quest for Intelligence)

Data Visualization of the Week

reddit/r/dataisbeautiful, sv-2, h/t Cecelia Dones (@DonesCecilia) from August 13, 2020



"The dataset used are metrics from Prometheus monitoring system." ... "This shows [1000+] time series plotted all at once in single graph."

Tools & Resources

AI Ethics Weekly newsletter

Lighthouse3 from July 27, 2020

"Start off your week right with our AI Ethics Weekly newsletter. Share this sign up link with your friends so they too can get this exclusive content delivered to their inbox every Monday!"

[Leveraging Machine Learning to Fuel New Discoveries with the arXiv Dataset](#)

Cornell University, *arXiv blog* from August 05, 2020

"To help make the **arXiv** more accessible, we present a free, open pipeline on **Kaggle** to the machine-readable arXiv dataset: a repository of 1.7 million articles, with relevant features such as article titles, authors, categories, abstracts, full text PDFs, and more."

Learning to Denoise Historical Music

arXiv, Electrical Engineering and Systems Science > Audio and Speech Processing; Yunpeng Li, Beat Gfeller, Marco Tagliasacchi, Dominik Roblek from August 05, 2020

"We propose an audio-to-audio neural network model that learns to denoise old music recordings. Our model internally converts its input into a time-frequency representation by means of a short-time Fourier transform (STFT), and processes the resulting complex spectrogram using a convolutional neural network. The network is trained with both reconstruction and adversarial objectives on a synthetic noisy music dataset, which is created by mixing clean music with real noise samples extracted from quiet segments of old recordings."

What Comes Next? Simple Practices to Improve Diversity in Science

ACS Central Science journal, Steven D. Townsend et al. from August 11, 2020

"In today's climate, a passionate plea against inequity (often delivered through social media or other powerful platforms) will rightfully and undoubtedly attract viewership. However, as the civil rights activists of the 1950s and 1960s would say, "What happens after the message is delivered?" In this editorial, we highlight examples of bias in science. Based on our experiences (chemists from industry, a historically black college and university (HBCU), and both private and public R1 universities), we offer solutions that will ensure scientists from underrepresented groups gain and maintain equal participation in science."

The Value in Science-Art Partnerships for Science Education and Science Communication

eNeuro journal, Cristian Zaelzer from July 02, 2020

"Just a fraction of the scientific knowledge produced in laboratories reaches a lay audience. Most of our communication with the public gets lost in translation because of the difficulties that science communication poses to scientists. Among other obstacles, differential exposure to scientific and critical thinking, discrepancies with social narratives, and communication training based in the deficit model add on top of a practice established on avoiding emotionality. In this context, effective communication requires the use of emotions, which are crucial to establishing trust. This commentary provides a rationale for collaboration with graphic design and fine arts to use emotions in science communication and education."

Computational Causal Inference at Netflix

Netflix Tech Blog, Jeffrey Wong and Colin McFarland from August 11, 2020

"Computational causal inference is a new, interdisciplinary field we are announcing because we want to build it collectively with the broader community of experimenters, researchers, and software engineers. The integration of causal inference into engineering systems can lead to large amounts of new innovation. Being an interdisciplinary field, it truly requires the community of local, domain experts to unite. We have released a whitepaper to begin the discussion. There, we describe the rising demand for scalable causal inference in research and in software engineering systems."

minGPT

GitHub - karpathy from August 17, 2020

"A PyTorch re-implementation of GPT training. minGPT tries to be small, clean, interpretable and educational, as most of the currently available ones are a bit sprawling. GPT is not a complicated model and this implementation is appropriately about 300 lines of code, including boilerplate and a totally unnecessary custom causal self-attention module."

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.

Data Science Community Newsletter Issue 203.