Data Science Community Newsletter features journalism, research papers and tools/ software for February 10, 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.

Welcome back to the newly remodeled Data Science Community Newsletter! As we entered 2021, our staff spent a little time revising the look of the DSCN and making some policy changes for non-news content (Jobs, Events, etc) in the interest of bringing you a more news-focused and overall shorter DSCN. Job announcements will be limited to academic tenure track or leadership positions only, and all jobs and events postings are limited to 1 line only. Please visit the ADSA Jobs Board and Events Calendar to find details and additional opportunities in and around our community.

We are also pleased to announce a new Sponsored Content box in the DSCN, offering our community the chance to highlight programs, events, and opportunities. We are grateful to the eScience Institute at the University of Washington for contributing the inaugural content. Contact info@academicdatascience.org to learn more about sponsored content.

Academic Data Science News

A new **National Bureau of Economic Research** working paper found that female academics, especially those who are parents, lost considerable time for their research. The work by **Tatyana Deryugina**, **Olga Shurchkov** and **Jenna E. Stearns** found that men and women without children also lost research time, but far less time than women

with children. And **Laura Lopez**, an **Ohio State** Astronomy professor, recently posted shockingly low levels of minority participation of minorities among Astrophysics PhD programs. Long term implications on academic diversity are to be determined, but these are setbacks that come at the same time as significant advances. Historically Black Colleges and Universities (HBCUs) have earned plaudits for student success despite less than ample resources. The point amplifies the tremendous potential payoff that **Apple** can expect from the company's $100 million commitment to create a learning hub across HBCUs for computer education. Also, **Northeastern University** appointed **Carla Brodley** to be the school's first Dean of Inclusive Computing, and **Elizabeth M. Renieris** is the founding director of the **Notre Dame-IBM Tech Ethics Lab**. **Cell Press** is inviting authors' submissions for how to better diversify journal authorship, and soon thereafter the journal *Cell* published the commentary, "Fund Black Scientists," authored by **Kelly Stevens** of **University of Washington** and a sizable network of women biomedical engineering faculty. These are impressive do-instead-of-say actions and a good model for institutions who are working on what they'll do. **Rachel Thomas** from **University of San Francisco** has good how-to advice for diversity initiatives. "First, what doesn't work: shallow, showy diversity efforts (even if they are well-intentioned) aren't just ineffective, they actively cause harm," she wrote in a recent tweet. Instead "spend time thinking through your strategy & making sure you can back it up."

"Chegging" is the new cheating. **Chegg**, an ed-tech company based in Santa Clara, offers a service to college students called Chegg Study which, for $14.95, lets them look up textbook and exam questions AND the answers to those questions. All of the questions (and answers) are right there in the Chegg database of 46 million solved problems, and students are turning in Chegg work as their own homework. Chegg also has a subsidiary career-planning service, **Thinkful**, which has partnered with **Arizona State** to create non-degree technical courses for bootcamp learning experiences. When a company has all of the problems and all of the answers, it's just packaging at that point.

Quick: what's the most commonly assigned college book in the US? Hint: you might call him the Original AI. Don't cheat. Or Chegg.

An analysis of advertised CS faculty searches for 2021 shows a 29% drop – not good, but not as bad as originally projected – whereas an overview of the state of data

science teaching and curriculum shows growth, but there remains a need for "consensus curriculum." **Boston University** advocates for a practicum, a classroom-industry hybrid, something that is at the center of the school's Masters in Statistical Practice program. Student journalists at **Stanford** investigating curriculum changes in the 2010s found that "the total number of computer science classes related to artificial intelligence more than tripled in the past decade from 25 classes to 77 classes." A new high school data science course has been approved by the **Arkansas State Board of Education**. Future curricula in any subject will rely on computer-based tools. **University of California-Berkeley**, **Google** and other researchers have gotten good results creating AI that can design lessons for itself, while a team at the **University of Michigan** has developed Mynerva, a cloud-based interactive textbook platform that adjusts automatically to the individual student. **Craig Smith's** *Eye on AI* podcast has been discussing AI in Edu-tech, most recently with the founders of **Riiid Labs**, a Silicon Valley startup.

The most commonly assigned college book in the U.S. is *Frankenstein* by **Mary Shelley**, as determined by **Degree Query**, a Utah ed-tech startup, using **Open Syllabus Project** data.

Nicole Hemsworth, reporter for *The Next Platform*, recently documented the computing hardware requirements of the **Institute for Health Metrics and Evaluation** (IHME) at the **University of Washington School of Medicine**. The Institute's beefed up infrastructure includes a 27,000-core computing cluster (and growing) on premises and four different **Qumulo** data appliances, all in the service of producing up-to-date models for global health populations. **David Lazer** and his group at **Northeastern University** also produce important ongoing COVID-19 research; their COVID States Project is regularly surveying thousands of Americans in all 50 states, and recently found that "there are still not enough people following public health recommendations to the extent necessary to stop the spread of COVID-19." The **Harvard Institute for Quantitative Social Science** (IQSS) is offering universities a turnkey app to dynamically model and simulate on-campus spread based on student interaction variables.

Static models are also impacting COVID therapies and outcomes. **Ohio State** professor **Ping Zang** created a new deep-learning model that can predict how human genes and medicines will interact. The model has identified at least 10

compounds that may hold promise as treatments for COVID-19. **NYU Langone Health**, working with **Facebook AI Research**, has a model that predicts in-hospital deterioration of COVID patients based on patient chest x-rays.

Twitter announced its Academic Research product track for the new Twitter API, allowing researchers access to what the company calls "the public conversation." **Facebook** says it will give researchers access to U.S. political ad data. These announcements come at a time when the **European Commission** is drafting the Digital Services Act, a pan-European set of rules for Internet companies to follow. Research access to social media data is a DSA rule, though the current draft has complicated data transparency rules, according to **Alex Engler** from the **Brookings Institution**. Adding to this mess: the eight-person European advisory for **Social Science One** resigned en masse in December. Social Science One is the academic data sharing organization set up by **Facebook** and led by **Gary King** and **Nathan Persily** that launched one year ago.

Collaboration is sometimes easier with hardware – supercomputers, robots, cars that are robots – and here the academic-industry romance is going strong. **NVIDIA** co-founder and **University of Florida** alum, **Chris Malachowsky**, has gifted his alma mater the HiPerGator AI supercomputer. **Graphcore** announced an academic program. The program will give free access and support for their systems; current partners include **University of California-Berkeley**, **Imperial College London**, **University of Bristol** (UK). **Purdue Polytechnic** is developing the first-of-its-kind smart manufacturing undergraduate program with the support of several industry partners, such as **Microsoft**, **Rockwell Automation**, **PTC**, and more. The **Toyota Research Institute** (TRI) added 13 more universities to participate in the next 5-year phase of its collaborative research program. The program is going after the hardest technical problems with self-driving cars, but without losing driving's human element. An example, TRI and **Stanford** have a self-drifting car, a souped-up Supra, in development to study the prevention of collisions.

Trust and security are crucial for distributed machine learning applications with human-computer interfaces. The **University of Southern California** and **Amazon** are launching a joint research center, The Center for Secure and Trusted Machine Learning, directed by **Salman Avestimehr**. Separately, the **Private AI Collaborative Research Institute** is a partnership led by **Intel** and

European cybersecurity companies, **Avast** and **Borsetta**. The group has enlisted its first set of funded academic projects to researchers at **Carnegie Mellon University**, **National University of Singapore**, **Technical University of Darmstadt**, **Universite Catholique de Louvain**, **University of California-San Diego**, **University of Southern California**, **University of Toronto**, **University of Waterloo** and **University of Wurzburg**.

Colby College in Maine kicked off a major education initiative with a $30 million gift from alumnus **Andrew Davis**, seeking to become the "First Liberal Arts College to Establish a Major Academic Program in Artificial Intelligence."

Rice University has a new Masters in Data Science degree program, available both online and on-campus.

Vanderbilt announced its new undergraduate data science minor starting Fall 2021.

Nazareth College in Rochester, New York, announced a new Institute for Technology, AI, and Society. ITAS will rollout its first minor — Business, Artificial Intelligence, and Innovation — in fall 2021. Three majors are planned: "Ethical Data Science," "Business, Artificial Intelligence, and Innovation" and "Technology, Artificial Intelligence, and Society."

Arkansas State University has been approved for a new Bachelors degree in Data Science and Data Analytics. One degree, two data.

University of San Diego (not **University of California-San Diego**, we've made that mistake before) hired **Maritza Johnson** to lead the school's new Center for Data, AI and Society.

Cal Poly in San Luis Obispo has pulled faculty from the university's six colleges in an effort to boost collaborative, data-driven, cross-disciplinary research and to "expand on-campus educational opportunities in the areas of data science, data analytics and data literacy."

Florida International University received $10.3 million from **NSF** to establish the Center for Internet Augmented Research for "network infrastructure and software

projects that will help scientists around the world share more data at a faster rate."

The **Canadian Institute for Advanced Research** now has more than 100 endowed CIFAR AI Chairs as the Pan-Canadian AI Strategy organization added 29 researchers to the community in January.

Fifteen leading research universities – **Brown**, **Caltech**, **Columbia**, **Cornell**, **Harvard**, the **University of Illinois**, **Michigan**, **Northwestern**, **Penn**, **Princeton**, **SUNY-Binghamton**, **UC Berkeley**, **UCLA**, the **University of Southern California**, and **Yale** – banded together and created a pooled technology licensing program, **University Technology Licensing Program LLC** (UTLP).

Editor's Picks

Journalist **Sharon Begley** died from lung cancer on January 16. Her final story is an expose on the 12% of lung cancer patients who never smoked. **STAT**, her publisher, noted, "She was a never-smoker."

The pre-Biden White House announced the new **National Artificial Intelligence Office**, tasked with implementing a national AI strategy under the leadership of Founding Director **Lynne Parker**, who also serves as U.S. Deputy Chief Technology Officer. (Journalist **Will Knight** described the seal of this new office as “an eagle wrestling a patriotic octopus inside a neural net.”) The **White House Office of Science and Technology Policy** established the AI Office in accordance with the National AI Initiative Act of 2020, which codified a number of policies and initiatives aimed at ensuring U.S. leadership in AI, which in turn was passed as part of the National Defense Authorization Act (NDAA) of 2021. The NDAA includes quite a few AI- and cybersecurity-driven initiatives for both military and non-military entities, including those related to ethics, bias, trustworthy AI and education and workforce training. Not to be outdone by **Department of Defense**, the **U.S. Department of Health and Human Services** (HHS) recently selected **Oki Mek** as its first-ever Chief AI Officer, and a new AI strategy outlines an approach to accelerate AI-centered pursuits across HHS. Over at the **Department of Education**, a letter was received from **U.S. Rep. Robert C. Scott**, a Virginia Democrat and the chairman of the **House Committee on Education and Labor** objecting to the use of student data by the **Pasco County (Florida) police department** to identify kids who might “fall into a life of crime.” Perhaps a patriotic octopus should pay a call on that police department?

The **U.S. Federal Trade Commission** showed regulatory teeth, pursuing (and settling) allegations that companies had misled consumers about their Internet and their health data. The enforcement arrives as personal tracking and profiling using data has become absolutely pervasive. **Veronica Barassi** describes how the profiling begins before birth in her **MIT Press** book excerpt. The stream of news reports on compromised privacy feels nonstop – "Period apps are a privacy nightmare," and "I looked at all the ways Microsoft Teams tracks users and my head is spinning." But technologies for rooting through personal data exhaust are also solving real problems – mental health and criminal investigation are examples. Cosmologist **Brian Nord** writes machine learning code; it's his job at **Fermilab** and **University of Chicago**.
In a **Gizmodo** interview he worries that his own algorithms might be biased against him,

with real life and work consequences, "I have to question whether I want to do AI work and how I'm going to do it; whether or not it's the right thing to do to build a certain algorithm. That's something I have to keep asking myself..."

Research News

Legacy lead piping for municipal water networks is a dangerous and costly problem. **Flint, Michigan**, has been forced to excavate more feet of neighborhood and household pipes than what's necessary to mitigate toxic lead leeching into city water. Machine learning to minimize the guesswork for legacy pipe excavation exists. It was used, then disregarded in Flint, but the technology developed first at **University of Michigan** and commercialized by **BlueConduit** is gaining adoption. Machine learning is essential for interpreting data from sensor networks to improve detection of pipe breaks and leaks where, at present, small fluid flow disruptions are impossible to detect reliably. **University of Pittsburgh** researchers are working on a nervous system for natural gas pipelines. German engineers at **RWTH Aachen University** can foresee reinforcing pipes with sensor-laden fabrics more durable than steel. Pipes' progress depends on successful machine learning, but eventually, and realistically, it should become a triumph of sustainability and effectiveness.

The Smart Cities research agenda is falling apart and thriving, both at the same time. **Cisco** recently dropped its Kinetic for Cities software in order to focus on Internet connectivity, addressing the "digital divide" that has become so starkly apparent as school and work moved home. The **Volkswagen Foundation**, responding to the global surge in commuting by car as people move away from mass transit, recently announced research grants including those to study the use of AI to reduce traffic congestion as well as to make parking more affordable.

NASA's Operation Icebridge mission to measure Arctic glacial ice thickness had its final radar-detector flight in December 2019. Nobody knows how much water is in these glaciers, an urgent problem given that the Arctic is warming at a faster rate than the rest of the planet. Researchers at the **University of Maryland, Baltimore County** are applying deep learning methods on the Snow Radar data to develop fully automatic methods to speed our understanding of polar ice sheets and snow thickness. The

research is described in the *Journal of Glaciology*.

In other news from our colder climes, advances in the isolation and sequencing of ancient DNA have begun to reveal the population histories of both people and dogs. Just-published research combines DNA data from ancient dogs and humans for evidence of the domestication of dogs by Siberian hunters over 23,000 years ago. The UK-US team analyzed previously sequenced mitochondrial genomes of more than 200 dogs from all over the world, some dating back 10,000 years.

The fast-growing field of pyroaerobiology seeks to study how microbial life is aerosolized and moved around by wildfire smoke. It's becoming increasingly important as wildfires become increasingly frequent. More drone research on aerosols, unmanned copters zoom over the Amazon rainforest in order to map and monitor the unique chemical signals, volatile organic compounds (VOCs), emitted by trees. Drone-based sensors collect data at vast scales, providing indications of how climate change, as well as human-caused deforestation and biomass burning, are leading to the potential release of massive pools of stored carbon. Further up, the Worldview-3 observation satellite provides high-resolution images that, coupled with deep learning algorithms, researchers are using to track the movements of African elephants, a species now classified as endangered with just 50,000 left in the wild.

University of Hawaii at Manoa oceanographers solved a controversy that has been debated in scientific literature for decades by fully reconciling climate and carbon cycle trends of the past 50 million years. Their research was published in *Science Advances*.

Technology worth transferring: **Texas A&M** researchers have invented algorithms to optimize traffic signal timing and reduce intersection waits.

Cornell is drilling an experimental borehole down to the Earth's crust, hoping to one day heat the snowy Ithaca campus geothermally.

Tweet of the Week

Twitter, Julie Kovacs Shane from Feb 4, 2021

Data Visualization of the Week

Twitter, Voilà: Francis Gagnon from Feb 7, 2021

Thread

Voilà: Francis Gagnon
@chezVoila

Today marks the 4th anniversary of @HansRosling's death. At @OutlierConf we had a little unconf discussion on Friday about what makes his BBC video so special.

Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Jo...
Subscribe and 🔔 to OFFICIAL BBC YouTube 👉 https://bit.ly/2IXqEInStream original BBC programmes FIRST on BBC
youtube.com

10:31 AM · Feb 7, 2021 · Twitter Web App

1 Retweet **3** Likes

Voilà: Francis Gagnon @chezVoila · 2h
Replying to @chezVoila
I had already covered what I see in a blog post but I learnt new things from the participants that I wanted discuss in this thread.

Events

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

Deadlines

Conferences

ACM CHI Workshop on Operationalizing Human-Centered Perspectives in Explainable AI

Online March 8-9. "We want to examine how human-centered perspectives in XAI can be operationalized at the conceptual, methodological, and technical levels towards a Human-Centered Explainable AI ." Deadline for submissions is February 14.

Contests/Award

Microsoft Research Faculty Fellowship Accepting Nominations

"Nominations are open for the 2021 **Microsoft Research** Faculty Fellowship through February 22. "

Studies/Surveys

We are embarking on a project to understand differences among Masters programs in Data Science.

As part of this project, we're interested in hearing what you think the top learning outcomes should be for a graduate of such a program. Let us know here: https://form.jotform.com/210105485654048

The Academic Data Science Alliance is gathering user stories about challenges faced in #DataScience publishing.

Hard to find a suitable venue for your code, toolchain, or workflow? Has your work been rejected because of a bad "fit"? Let us know!

Tools & Resources

[P] Connected Papers partners with arXiv and Papers with Code

reddit/r/MachineLearning, Discordy from Feb 5, 2021

"We have partnered with arXiv.org and from now on every paper page in arXiv will link to a corresponding Connected Papers graph."

The Montreal AI Ethics Institute has published "The State of AI Ethics Report"

Twitter, The Institute for Ethical AI & Machine Learning from Feb 6, 2021

It "captures the most relevant developments in AI Ethics since October of 2020."

Doing it Right: A Better Approach for Software & Data

Dryad blog, daniellalowenberg from Feb 8, 2021

"The Dryad and Zenodo teams are proud to announce the launch of our first formal integration. "

Careers

See the ADSA Jobs Page for more opportunities

About Us: The Data Science Community Newsletter was founded in 2015 in the Moore-Sloan Data Science Environment at NYU's Center for Data Science. We continue to be supported by the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation through the Academic Data Science Alliance. The newsletter is written and the content is compiled by the Academic Data Science Alliance. Our archive of newsletters is at cds.nyu.edu/newsletter and is the process of transitioning to another, permanent location. Our mailing address is1037 NE 65th St #316; Seattle, WA 98115.

OPT OUT: If you do not want to receive this newsletter, please email brad.stenger@gmail.com with the word 'unsubscribe' in the subject line.

OPT IN: Feel free to forward the Data Science newsletter to colleagues. They can sign up for the newsletter using this web form.