

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for September 23, 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](#).

RIGHT TO REPAIR GAINS MOMENTUM

Legislative momentum is increasing from farmers, soldiers, the open source community, and independent repair shops for a Right to Repair devices and machines to make them easier and cheaper to repair, less likely to be tossed in a heap of e-waste. Device manufacturers from the technology, agriculture, and medical device industries oppose expanded consumer rights. [Click through](#) to get the full story in our latest *From the Desk of Laura Norén*.

STANFORD'S PRE-PRINT ON FOUNDATIONAL MODELS

A recent [pre-print](#) co-authored by 32 faculty and 117 research scientists, postdocs, and students at **Stanford University** proposed that the field of artificial intelligence is still developing, but sufficiently mature that it is possible to identify the key foundational models from which many applications of artificial intelligence be tuned to fit. The paper was not received well by **Gary Marcus** and **Ernest Davis**, who [argue](#) that "there is no good reason, beyond mere hope, to think that Foundation models" would be able to extract meaning from a corpus of text and take an appropriate action or perform other human-like reasoning. **Mark Reidl** a **Georgia Tech** linguist, threatened "[violent eye-rolling!](#)" at any reference to foundational models, indicating he feels the argument, "is a brilliant and disgusting PR stunt." **Emily Bender**, also a linguist with a strong command of Twitter, [tweeted](#) that "overstating" what LLMs can do — that they are able to understand natural language and participate in open-ended conversation (which they don't and can't), presents significant roadblocks to obtaining funding for research into other natural language training sets and models. As challenging as the past 18 months have been, it's... nice to be contemplating a heated debate about data science that has nothing to do with politics or existential demise.

Featured Jobs

See the [ADSA Jobs Page](#) for more opportunities.



[Data Scientist at Acorn AI](#)

Location: New York, Boston

LAWS WE LIKE

The UK has [announced it will rewrite part of its legislation](#) to impose stiffer fines and penalties on parties responsible for spam calls and texts. Britain's **Information Commissioner's Office** will also examine "what more can be done to mitigate algorithmic bias". With rather grandiose optimism, the Digital Secretary **Oliver Dowden** exclaimed that the changes are expected to usher in "a new golden age of growth and innovation right across the UK".

THE ENERGY GRID OF THE FUTURE TODAY NEEDS NETWORK SCIENCE

Of the fourteen deaths in greater New Orleans attributed to Hurricane Ida, nine were caused by "excessive heat during an extended power outage" according to a [recent report](#) by **ProPublica** and **NPR**. Transitioning away from an inadequate electricity grid will [require substantial network analysis and research](#), one item in a fifteen-item list, part of a white paper authored by **BloombergNEF**, **World Economic Forum** and **Deutsche Energie-Agentur** (Germany's energy agency). "Energy transition networks at a scale of billions of things will be too big to be run only by humans. They will require artificial intelligence," says **Nathaniel Bullard**, BloombergNEF's Chief Content Officer. This is particularly true as the network becomes more complex with more nodes acting as generators *and* consumers *and* battery storage.

Because the US national grid is so interconnected, disasters and peaks somewhere will impact people across the country as will the lack of accurate peak prediction, pricing, and storage utilization. If anyone knows of a national effort to upgrade the entire country to a smarter, more renewable-accommodating, peak-shaving grid, please send us tips and links!

WILDFIRE

Andrew Frickers almost lost his home in the 2018 Camp Fire in Butte County, California. A computational social scientist at **Cal Poly San Luis Obispo**, Frickers has put three years into [developing "Damage Map."](#) an AI-informed, quickly-assembled wildfire destruction zone pre-report. Frickers' Damage Map proved 90 percent accurate for 2021 wildfires and should become more accurate with additional training. Wildfire prediction models used for catastrophe insurance are far more complex. The new [U.S. Wildfire Model Version 1.0](#) by **Karen Clark & Co.** incorporates nine different climate change scenarios based in part on the **Intergovernmental Panel on Climate Change** report. Bottom line impacts for insurers will exceed \$10 billion, possibly \$30 billion, according to Ms. Clark. To mitigate destruction localities are increasingly [turning goats loose](#) to graze on brush and weeds, and reduce wildfire fuel. To speed claims processing in the d-

zone **Farmers Insurance** will be using [custom-built four-legged robots](#) created by **Boston Dynamics**. It's inevitable: Robot goats. And \$50 billion cat insurance models.

SPONSORED CONTENT



Become a Founding Member of ADSA!

There's still time left, join by **October 1st** for special institutional member benefits

<https://academicdatascience.org/get-involved/adsa-membership>

If your institution pays membership dues by October 1st, 2021, we will list you as a Founding Member organization on our website and in our Annual Member Book. Your institution will also receive 1 FREE Content Box in the DSCN. And we will give you a podium shout-out at the 2021 Leadership Summit and Annual Meeting! Click to [Learn More](#).

AGRICULTURAL TECHNOLOGY

Our readers are highly clustered in urban areas where there's little farming. But everyone eats and the way we eat impacts our environment, our climate future, and public health. Feeding large groups of people is one of those hard problems that goes back to Biblical times (see: loaves, fishes) and has typically been driven by goals to increase yield, reduce waste, and please finicky palates. The green revolution certainly delivered — yields went way up for many crops. Environmental and public health impacts have not always been widely considered, leading to situations such as overuse of water resources and underutilization of efforts to contain negative externalities — animal waste products, synthetic fertilizers, pesticides, herbicides. Mapping and manipulating plant genomes led to hardier plants as well as legal precedents that allowed corporations to patent genes.

It is against this backdrop that we present a range of data science and AI applications in agriculture.

At the **University of Florida**, Professors **Raluca Mateescu** and **Fernanda Rezende** have genetically sequenced about one thousand cows and are using a supercomputer to [predict which genes are linked to higher meat and milk yields](#). If their model works as they hope it will, they'll be able to advise cattle breeders. However, the **Alexandre Family Farm**, a large dairy farm in California, has [created a more succinct dairy cattle ecosystem](#) by planting grasses their cows eat while fertilizing on a monthly rotating schedule. With this technique, no waste lagoons, little non-grass feed, nor any advanced data science is required to increase the farm's efficiency and reduce its negative externalities.

Nurturing life at scale, in a monoculture unlike what nature would have provided, is hard. Deriving meaning from, say, various types of humidity, temperature, and nutrient level sensors all over a greenhouse and turning them into appropriate action has turned out to be a fruitful venture. (Pun intended!) The third year of the [Autonomous Greenhouse Challenge](#) has once again shown that it is possible for AI-assisted greenhouses to outperform humans, improving both yield and profit (which are, of course, decidedly human goals that nature may or may not co-sign).

The **National Science Foundation** and the **US Department of Agriculture** joined forces to [fund 11 new AgAI research institutes](#), funding \$140 million worth of research in 2020 and committing \$220 million to grants in 2021. It's not clear if the funding for the new \$25 million Center for Research on Programmable Plant Systems (CROPPS) at **Cornell University** backed by the **NSF** came out of this bucket. What is important is for the mostly city-dwelling readers of the DSCN to realize how important AgAI is and isn't. The history of inserting technology into the food production cycle is long, largely beneficial to humans, but not necessarily all that great for the overall sustainability of life on this planet. Be careful what you optimize. If it's only crop yield and profit, please reconceive.

DEEP(LY NERDY) IMPORTANT ADVANCES / HARDWARE

MIT, Boston University, and Maynooth University in Ireland "created the first silicon chip that is [able to decode any code](#), regardless of its structure, with maximum accuracy, using a universal decoding algorithm called Guessing Random Additive Noise Decoding (GRAND). By eliminating the need for multiple, computationally complex decoders, GRAND enables increased efficiency that could have applications in augmented and virtual reality, gaming, 5G networks, and connected devices that rely on processing a high volume of data with minimal delay." These advances in hardware often precede significant efficiency gains in commercial applications and the (continuing) proliferation of software in every imaginable context. In this particular example, I would imagine this research will be particularly useful for edge computing which is likely to be a crossroads teaming with IoT gadgetry, 5G signals, and security threats/bandaids of all sorts. With that much going on, a bit of streamlining would be useful.

In an even deeper investigation into the nature of the crystalline structure of silicon, a group of international scientists led by a team at **NIST** has [developed an extremely precise technique for measuring atomic and subatomic characteristics](#). Not only is their work important for potential commercial applications, it could advance research on the "fifth force", which would advance basic scientific understanding of atomic behavior.

FOLLOW THE MONEY

\$2,100,000,000 **Joe Biden and Kamala Harris Administration, American Rescue Plan fund** - > **Centers for Disease Control** to ["strengthen and equip state, local, and territorial public health departments...to better fight infections in US healthcare facilities](#), including COVID-19 and other known and emerging infectious diseases" such as MRSA.

\$470,000,000 **National Institutes of Health** -> **NYU Langone Health** to create the [Researching COVID to Enhance Recovery \(RECOVER\) Initiative](#) which will study long COVID, beginning by enrolling 30-40,000 volunteers as they are diagnosed with COVID-19.

\$225,000,000 **John Morgridge (\$75m committed with another \$50m match available)**, the **Wisconsin Alumni Research Foundation** and as-yet unidentified future donors (\$50m) - > **University of Wisconsin-Madison** for the new [School of Computer, Data & Information Sciences](#) facility which is slated to open in 2024 and will house data science, information science, and statistics. No state funding is being used.

\$61,700,000 **National Institutes of Health** -> **Columbia University, Vagelos College of Physicians and Surgeons** to "expand training opportunities and offer a ['data concierge service'](#) that will link researchers with experts in biostatistics, biomedical informatics, and data science. They will also engage community leaders as 'ambassadors' to promote research that improves community health."

\$50,000,000 **National Science Foundation** Convergence Accelerator program -> 10 research teams to create [“practical, sustainable applications in quantum technology, combating climate change and managing natural disasters, monitoring and detecting biothreats”](#)

\$37,000,000 **U.S. Department of Energy** -> 14 awardees [Artificial Intelligence and Machine Learning at DOE Scientific User Facilities](#)

\$28,000,000 **U.S. Department of Energy** -> 5 awardees [4 in California, 1 in Florida; all PIs are men] for [Advanced Computational Research in the Sciences](#)

\$26,000,000 **U.S. Department of Energy** -> 10 awardees to [Advance Chemical and Materials Sciences with Data Science](#)

\$25,000,000 **Jim Goetz** -> **University of Cincinnati** to [rename a building](#) after beloved professor **Thomas D. Mantei**. More fitting and sustainable than naming after the donor, right?

\$8,000,000 **National Science Foundation** -> **University of Southern California, University of North Carolina, Indiana University, University of Notre Dame, University of Utah, and Texas Tech** for a [pilot](#) Cyberinfrastructure Center of Excellence with a focus on data lifecycle management.

\$4,400,000 **National Science Foundation** -> **University of Massachusetts Amherst** in a renewal of [its CyberCorps Scholarship](#) for Service (SFS) program

\$3,000,000 **National Science Foundation** -> **University of Texas at Austin** [“Privacy-Preserving Abuse Prevention for Encrypted Communications Platforms”](#) to help prevent abuse on encrypted messaging apps like WhatsApp and Signal while preserving guarantees of private communications

\$2,999,513 **National Science Foundation** NRT -> **Stony Brook University** “to [provide interdisciplinary cross-training](#) to PhD students in the data sciences alongside PhD students in the human-centered sciences to detect and address biases in data, models, people and institutions.”

\$2,900,000 **University of California, Office of the President** -> **UCLA Extension** to [offer low- and no-cost extension certificates](#) in substance abuse counseling, data science and early childhood education to underrepresented and underserved communities.

\$2,000,000 **National Science Foundation** -> **University of California-Berkeley** for the [Effective Programming, Interaction, and Computation with Data \(EPIC\) Lab](#) “to improve the usability of big criminal justice datasets for public defenders and others”

\$1,800,000 **National Science Foundation** -> **Washington University in St. Louis** to [“conduct large-scale simulations to identify candidates for artificial multiferroics, a new type of quantum material.](#)...that integrate two-dimensional van der Waals magnets with ferroelectrics.”

\$1,600,000 **Gordon and Betty Moore Foundation** -> **University of Rochester** to [“create viable superconducting materials”](#) at room temperature.

\$1,000,000 **Department of Energy** -> **National Institutes of Health** grantees at **Argonne National Laboratory** with **Lawrence Livermore National Laboratory**, the **University of Chicago**, the **Broad Institute**, and **Massachusetts General Hospital** for [PALISADE-X: Privacy Preserving Analysis and Learning in Secure and Distributed Enclaves and Exascale Systems](#). Spread across this many institutions with a problem that is not solved, doesn't seem like \$1 million is enough.

\$55,000 (times 20) **Research Corporation for Science Advancement + Heising-Simons Foundation + NASA + The Kavli Foundation** -> 20 Scialog Fellows on 8 teams [to advance “our understanding of the habitability of planets](#), of how the occurrence of life alters planets and leaves signatures, and of how to detect such signatures beyond Earth, in our solar system and on exoplanets”.

NEW PROGRAMS

National Science Foundation [announced](#) 6 new Science and Technology Centers:

- * NSF Center for [Chemical Currencies of a Microbial Planet](#)
- * NSF Center for [Integration of Modern Optoelectronic Materials on Demand](#)
- * NSF Center for [Learning the Earth with Artificial Intelligence and Physics](#)
- * NSF Center for [Oldest Ice Exploration](#)
- * NSF Science and Technologies for [Phosphorus Sustainability Center](#)
- * NSF Center for Research on [Programmable Plant Systems](#)

[National AI Advisory Council](#) **U.S. Department of Commerce** which will “issue recommendations on U.S. AI competitiveness, workforce equity, funding, research and development, international cooperation, and legal issues” and is now recruiting.

[AI and Law Enforcement subcommittee](#) **U.S. Department of Commerce** to focus specifically on AI and justice.

[Raven Precision Agriculture Center](#) **South Dakota State University**. The facility cost \$46.1m; the grand opening was celebrated with speakers, tours, and “the Precision Agriculture Bowl football game”.

[Master of Science in Urban Analytics](#) **Georgia Tech** which has an extremely ambitious goal of fixing “big city problems” with the fields of “urban planning, computing, and industrial and systems engineering” though curiously not with sociology, architecture, economics, business, or earth/atmospheric science.


[Data Science Minor](#) **University of British Columbia**, a joint offering between the Departments of Computer Science and Statistics.

[First Year Focus: Computation](#) University of British Columbia in which UBC Science frosh to opt into a cohort with which they “take a common set of five online computer science, math, data science and communications courses”.

DATA VISUALIZATION OF THE WEEK

Twitter, Jonathan McKinney, from September 18, 2021

← Thread

 Jonathan McKinney (空歩)
@mc_kupo

Have you seen this graphic in [@Abebab et al 2021](#)!?! It provides a striking look into the values of ML/AI research. Notice the red and blue text. That's where human social / ethical concerns begin. This meta-analysis is worth meditating on. [#digitalhumanities](#) should pick this up.



Figure A.1: Value occurrences across papers.

1:24 PM · Sep 18, 2021 · Twitter Web App

And the [follow-up analysis](#) by Thomas Dieterich of Oregon State University is something we wish we'd written for DSCN.

Deadlines

Studies/Surveys

[Do you have an ideas for people to help organize #FAccT2022 or later installments?](#)

"We want org teams to reflect the diversity of the community & give new community members opportunity to contribute. Make your suggestions here (self-nominations welcome)"

Conferences

[Ninth international congress on peer review and scientific publication—call for abstracts](#)

Chicago September 8-10, 2022. Deadline for abstracts submissions is January 31, 2022.

Education Opportunities

[Second annual Rising Stars in Data Science workshop at University of Chicago](#)

Chicago November 11-12. "The Rising Stars workshop focuses on celebrating and fast-tracking the careers of exceptional data scientists before they transition to postdoctoral scholar, research scientist, industry research position, or tenure track position." Deadline for applications is September 30. Deadline for faculty nominations is October 4. [Virtual information session](#) is September 24 starting at 3 p.m. Eastern.

Tools & Resources

[Simulations are about to get way, way faster with JuliaSim](#)

Not a Monad Tutorial blog from September 2, 2021

"JuliaSim is a cloud-based simulation platform built on top of the Julia open source stack, including SciML and ModelingToolkit "

[Race and Data Science Resources](#)

Columbia University, Data Science Institute

A long list. Necessarily so.

Events

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

[Academic Data Science Alliance, Open Research & Open Scholarship + Ethics SIG](#)

"This Thursday Sept 23 at 10 Pacific, join our conversation with **Dennis Wei** at **IBM** about the AI Fairness 360 Toolkit and how you can use it to examine, report, and mitigate bias in ML models throughout the AI application lifecycle." [free, registration required]

NYU NLP & Text-as-Data Speaker Series Returns

New York City September 23, starting at 4 p.m. Eastern, with speaker **Ankur Parikh** from **Google**.

[#GovDATAx 2021 is airing Sept. 29-30!](#)

Online September 29-30. Speakers include: @RepDonBeyer, @SpeakerRyan, @fema's **Melissa Forbes**, @HHS_ASPE's **Naomi Goldstein**, @DeptVetAffairs's **Kshemendra Paul**, @CommerceGov's **Natalie Evans Harris**, & more!

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](#). Our archive of newsletters is at <https://academicdatascience.org/resources/newsletter>. Our mailing address is [1037 NE 65th St #316; Seattle, WA 98115](#).