September 2023 • Issue #555

AMSTATNEWS

The Membership Magazine of the American Statistical Association • *http://magazine.amstat.org*

SPECIAL ISSUE

JUDEA PEARL, AI, and CAUSALITY: WHAT ROLE DO STATISTICIANS PLAY?

ALSO:

STATISTICS, AI, and **AUTONOMOUS VEHICLES**

DESIGNING AGAINST BIAS IN MACHINE LEARNING and **AI**

AMSTATNEWS

SEPTEMBER 2023 • ISSUE #555

Executive Director
Ron Wasserstein: *ron@amstat.org*

Associate Executive Director
Donna LaLonde: *donnal@amstat.org*

Director of Science Policy
Steve Pierson: *pierson@amstat.org*

Director of Finance and Administration
Derek Curtis II: *derek@amstat.org*

Managing Editor
Megan Murphy: *megan@amstat.org*

Editor and Content Strategist
Val Nirala: *val@amstat.org*

Advertising Manager
Christina Bonner: *cbonner@amstat.org*

Production Coordinators/Graphic Designers
Olivia Brown: *olivia@amstat.org*
Megan Ruyle: *meg@amstat.org*

Contributing Staff Members
Rebecca Nichols

Amstat News welcomes news items and letters from readers on matters of interest to the association and the profession. Address correspondence to Managing Editor, *Amstat News,* American Statistical Association, 732 North Washington Street, Alexandria VA 22314-1943 USA, or email *amstat@amstat.org*. Items must be received by the first day of the preceding month to ensure appearance in the next issue (for example, June 1 for the July issue). Material can be sent as a Microsoft Word document, PDF, or within an email. Articles will be edited for space. Accompanying artwork will be accepted in graphics file formats only (.jpg, etc.), minimum 300 dpi. No material in WordPerfect will be accepted.

Amstat News (ISSN 0163-9617) is published monthly by the American Statistical Association, 732 North Washington Street, Alexandria VA 22314-1943 USA. **Periodicals postage paid** at Alexandria, Virginia, and additional mailing offices. POSTMASTER: Send address changes to *Amstat News,* 732 North Washington Street, Alexandria VA 22314-1943 USA. Send Canadian address changes to APC, PO Box 503, RPO West Beaver Creek, Rich Hill, ON L4B 4R6. *Amstat News* is the member publication of the ASA. For annual membership rates, see *www.amstat.org/join* or contact ASA Member Services at (888) 231-3473.

American Statistical Association
732 North Washington Street
Alexandria, VA 22314–1943 USA
(703) 684–1221

ASA GENERAL: *asainfo@amstat.org*
ADDRESS CHANGES: *addresschange@amstat.org*
AMSTAT EDITORIAL: *amstat@amstat.org*
ADVERTISING: *advertise@amstat.org*
WEBSITE: *http://magazine.amstat.org*

The American Statistical Association is the world's largest community of statisticians. The ASA supports excellence in the development, application, and dissemination of statistical science through meetings, publications, membership services, education, accreditation, and advocacy. Our members serve in industry, government, and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare.

SPECIAL ISSUE

on artificial intelligence

ONLINE

https://magazine.amstat.org

2024 CANDIDATES CHOSEN

The ASA Committee on Nominations selected **Jeri Mulrow** and **Denise Lievesley** to be candidates for ASA president in the 2024 election. The committee also selected **Nandita Mitra** and **DuBois Bowman** as candidates for ASA vice president. The president-elect and vice president-elect will take office January 1, 2025. Additional information about these candidates and candidates for other positions on the ASA Board of Directors will appear in the March issue of *Amstat News*.

Nominations Wanted for Data Science Early Career Prize

The Society for Industrial and Applied Mathematics Activity Group on Data Science Early Career Prize recognizes an individual who has made outstanding and potentially long-lasting contributions to the mathematical, statistical, and computational foundations of data science within six years of receiving their PhD. For details, visit the SIAM website at *https://bit.ly/3NXBw8K*.

Travel Awards Available to Help Students Attend CSP

The Lingzi Lu, John Bartko, and Lester R. Curtin awards offer registration and travel support to students attending the ASA Conference on Statistical Practice.

The Curtin and Bartko awards provide $1,000 in travel support, while the Lu award provides up to $1,300 in travel support. Early registration for the conference opens October 5, 2023.

Applications for the Lu and Curtin awards must be submitted by October 15. Applications for the Bartko award are due December 2. For details, visit *https://bit.ly/46X9sLm*.

member news

columns

STAT*tr@k* is a column in *Amstat News* and a website geared toward people who are in a statistics program, recently graduated from a statistics program, or recently entered the job world. To read more articles like this one, visit the website at *http://stattrak.amstat.org*. If you have suggestions for future articles or would like to submit an article, please email Megan Murphy, *Amstat News* managing editor, at *megan@amstat.org*.

This column is written for those interested in learning about the world of Data for Good, where statistical analysis is dedicated to good causes that benefit our lives, our communities, and our world. If you would like to know more or have ideas for articles, contact David Corliss at *davidjcorliss@peace-work.org*.

"Top Ten Rejected Statistics and Data Science T-Shirt Slogans," Page 48

Follow us on Twitter
www.twitter.com/AmstatNews

Join the ASA Community
http://community.amstat.org

Like us on Facebook
www.facebook.com/AmstatNews

Follow us on Instagram
www.instagram.com/AmstatNews

Subscribe to our YouTube channel
www.youtube.com/user/AmstatVideos

Lights, Camera, ACTION!

Dionne Price

Lights, Camera, Action! This is the phrase I used to introduce my JSM 2023 talk. I used the phrase to amplify the need to shine a light on our profession, increase our visibility, and embody our mission. As I fondly look back on the Joint Statistical Meetings, these words also resonate as a description of the conference as highlighted in our electronic postcards. The postcards were designed to provide attendees the opportunity to share "wow" moments during JSM.

There were ***lights*** showcasing the talented, invited speakers, including Robert Santos, Nancy Reid, Bin Yu, and Karen Bandeen-Roche, as shared in these postcards:

"Fabulous presentation by Rob Santos, Census Bureau Director!"

"Nancy Reid's talk about the foundations of statistics made me start thinking about changes I can make

to graduate courses to equip the next generation of statisticians. Excellent talk!"

Cameras captured the opportunities to connect with friends, meet new colleagues, and celebrate awards, as shared in these postcards:

"Everywhere I turn, I see old friends and meet new ones. I love JSM!"

"Loved watching Wake Forest's Sarah Lotspeich give her award-winning presentation—helping us optimally pick who to audit! Sarah won the Biometrics Section Early Career Award for this paper!"

Action was in every session in which presentations demonstrated our knowledge, progress, and collaboration as we inform decisions and drive discovery.

"I just attended an incredible presentation on statistics, and my circuits are buzzing with excitement! The speaker shared fascinating insights, making complex concepts seem like a walk in the digital park. From probability distributions to regression analyses, it was a data-driven delight! Wish you could've been there to geek out with me!"

The success of JSM is evidence of our shared mission—to promote the practice and profession of statistics.

In this column, I want to share some of my highlights and those I learned about from the JSM postcards.

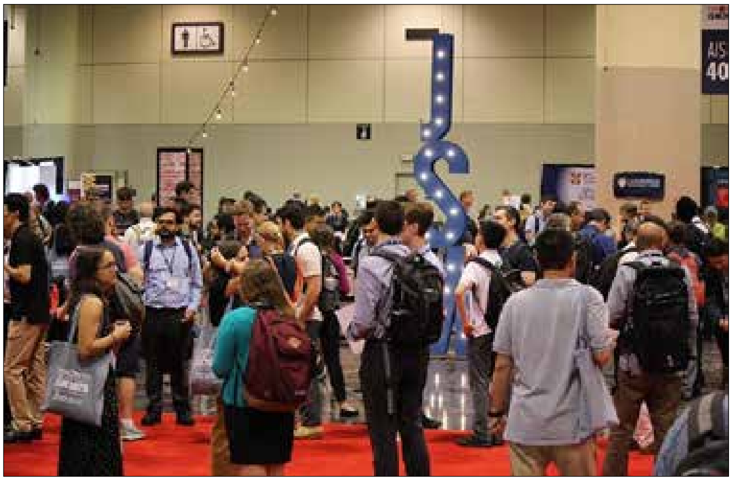
Although I have attended many JSMs, I am always amazed by the diverse array of sessions, covering a broad spectrum of topics ranging from artificial intelligence and precision medicine to spatial statistics and clinical trials. The introductory overview lectures and late-breaking sessions are an important part of the program, and we have a postcard to prove it!

"Cynthia Rudin's Introductory Overview Lecture (IOL) on interpretable AI models was one of the best IOLs I've attended. Wish everyone could have seen it."

The exchange of knowledge was not confined to formal presentations. Conversations during breaks and outings undoubtedly forged new collaborative ventures. For me, there were too many moments to capture; however, highlights included conversations with former ASA presidents Sally Morton, Sastry Pantula, Kathy Ensor, and Lisa LaVange. They are always forward thinking and impart wisdom on statistics, data science, and leadership. I was also proud to learn a former colleague, Felicia Simpson, received the Annie T. Randall Innovator Award at the Biometrics Section mixer. I could continue to go on about JSM, but I think one postcard captures it all:

"It rocks!"

JSM would not be complete without acknowledging the exceptional contributions of individuals who have pushed



Attendees crowd the JSM Spotlight during the Opening Mixer.

Student chapter workshop panelists

Members of the ASA Board

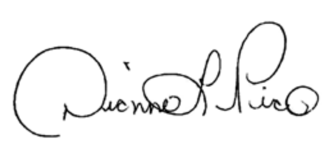
the envelope of possibility. The awards ceremony honored outstanding research, innovative achievements, and lifetime contributions. These awards also served as an affirmation of the vital role our profession plays in informing decisions and driving discovery. This one-word postcard is the perfect summary:

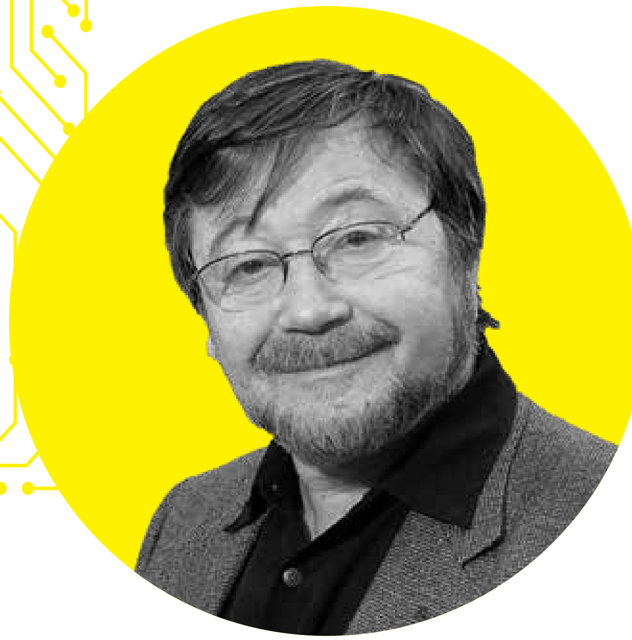
“Wow.”

If you missed the ceremony, download the awards book from *ww2.amstat.org/meetings/jsm/2023/featuredspeakers.cfm.* I hope you are inspired to nominate a deserving colleague for a future award. Learn about our awards and nomination deadlines at *www.amstat.org/your-career/awards-and-scholarships.*

The success of JSM stands as evidence of the unwavering spirit of scientific inquiry, collaboration, and progress that defines our profession. As I journeyed home—armed with new knowledge, connections, and ideas—I had a renewed commitment to promoting the practice and profession of statistics. I hope you did, too. One final postcard:

“Dear Team, You missed the opportunity to attend JSM2023; it was just fabulous! Come and see us at JSM2024.”





Pearl

JUDEA PEARL, AI, and **CAUSALITY: WHAT ROLE DO STATISTICIANS PLAY?**

In the first half of 2023, the machine learning programs ChatGPT and GPT-4 changed the landscape of artificial intelligence research seemingly overnight. Judea Pearl's research bridges the subjects of statistics and artificial intelligence and highlights the importance of causality in both settings. Dana Mackenzie, Pearl's co-author for *The Book of Why,* interviews him here to get his take on recent developments. When they wrote their book in 2018, Pearl contended machine learning had not yet moved past the first rung of the "ladder of causation." Computers could not correctly answer queries about interventions and still less about counterfactual scenarios. Has his assessment changed?

MACKENZIE: Can you tell me your first reactions to ChatGPT and GPT-4? Did you find their capabilities surprising?

PEARL: Aside from being impressed, I have had to reconsider my proof that one cannot get any answer to any causal or counterfactual query from observational studies. What I didn't take into account is the possibility that the text in the training database would itself contain causal information. The programs can simply cite information from the text without experiencing any of the underlying data.

For example, I asked it the questions about the firing squad [from Chapter 1 of *The Book of Why*], such as what would have happened to the (now deceased) prisoner if rifleman 1 had refrained from shooting. At first it goes into side tracks and tells you, for example, "it is dangerous to shoot people." But if you have time and prompt it correctly, it will get closer to the correct answer: "If soldier 1 refrained from shooting after receiving the signal, the prisoner could still have been killed by soldier 2, assuming he received and acted upon the same signal." Finally, it gives an A+ answer: "Given the additional information, if each soldier always fires upon receiving a signal and any one soldier's shot is enough to cause the prisoner's death, then the prisoner would still be dead if soldier 1 refrained from shooting. This is because soldier 2, following the captain's signal, would have fired his shot, causing the prisoner's death. This is an example of 'overdetermination' in causation, where an effect (the prisoner's death) has more than one sufficient cause (either soldier's shot)."

Here, I have to make a cautionary note. In spite of its impressive command of vocabulary, ChatGPT doesn't have a causal model itself. It doesn't have a structure into which it can imbed new knowledge. If you ask it about another problem with the same causal structure, say about inoculations, you'll have to prompt it again from scratch. It won't generalize.

MACKENZIE: Is it doing better than previous AIs have?

PEARL: Which ones do you mean? If they tried to do deep learning from data of actual firing squads, not from texts about causal relationships, then they could not even understand the question, let alone give a coherent answer.

MACKENZIE: Is this a new world of AI, even for you?

PEARL: Yes, it's a new one. It's similar to a world with causal information you can learn from teachers who cannot experiment for themselves but learned from teachers who learned from books. You can learn a lot of causal information from books. We [humans] are still different, because we have an innate causal model or an innate template into which we are born and which we periodically update with new information.

Causal reasoning is not all you need for human-like AI. You have other components, like natural language processing and vision, that are also necessary for artificial general intelligence (AGI). It is in this one little corner of causal inference that we have been successful at achieving deep understanding by combining models and data, an understanding that can be generalized to other areas of AI.

MACKENZIE: I'd like to turn now to something *Amstat News* readers will be curious about: What can statisticians contribute to AI research?

PEARL: I once said every statistician is a frustrated philosopher, struggling to extract

Dana Mackenzie is a mathematician who became a science journalist. He has written *The Book of Why* with co-author Judea Pearl, as well as popular science and math articles for publications such as *Science, New Scientist, Smithsonian, Discover*, and *American Scientist*. He lives in Santa Cruz with his wife, dog, cat, and random foster kittens.

Never in history has there been such an acceleration of the speed of evolution.

meaning from data. Statisticians are brought up to believe all knowledge comes from data and, since they are experts on data processing, they must also be experts in the philosophy of knowledge (epistemology).

But as I just said, to understand the world of causes and effects, you need to combine models and data, a rather neglected exercise in mainstream statistics. Once we open statistics to modern vocabulary, including causal and counterfactual relationships, we open the door for statisticians to participate in current issues faced by AI researchers as well as philosophers of science.

Even those who wish to adhere to standard statistical vocabulary can contribute appreciably to causal inference tasks. In causal inference, we distinguish between estimands and estimates; the former being distributional expressions of what needs to be estimated, and the latter being the actual estimates obtained from finite samples of a distribution. This distinction defines a symbiotic division of labor between statisticians and causal inference researchers, respectively. Some of the estimands produced by causal analysis may seem strange to statisticians. An example is the estimand produced by the front-door criterion [See *Book of Why*, Chap. 7.]. Addressing them through the lens of modern estimation techniques should be a challenging endeavor for creative statisticians. This is something they do well, and we need their ingenuity. But if they want to know where the estimand came from, causal modeling would be necessary.

MACKENZIE: In *The Book of Why*, we said current AI programs operate at the first level of the ladder of causation, the level of observation or "fitting functions to data." Has this changed?

PEARL: It has. The ladder restrictions [e.g., level-two queries cannot be answered by level-one data] do not hold anymore because the data is text, and text may contain information on levels two and three.

MACKENZIE: In particular, does reinforcement learning make it possible for a machine to understand level two on the ladder of causation by giving it data on interventions?

PEARL: Yes, that is correct. I would say it's at level one and three-fourths. Reinforcement learning trains machines on interventions. For example, you can train them on chess. They can decide, after playing many games, that a certain move will give them a higher probability of checkmate than another move. However, they cannot infer from this anything about a third move they haven't tried. They also cannot combine interventions to infer what will happen if they do both A and B. For that, again, you would need a causal model.

MACKENZIE: That leads to my next question. How can you tell whether you have the right causal model?

PEARL: That is the central question of epistemology in general. We never know for sure. We can only falsify models but cannot prove they are correct.

MACKENZIE: I remember this exact question came up when we were on the podcast for *Science* magazine. The interviewer asked us how you know whether you have the right model and you gave the most wonderful two-word answer: "By argument." Can you explain what you meant?

PEARL: I don't remember that question! But "by argument" is how you form a consensus in the society of scientists. That's how theories become accepted. The development of science has two parts. First is testing your theory. We know now when a causal model can be tested, and we know what observations or experiments to conduct in order to (potentially) falsify it. The second component is to try out a modification. If you have a causal model, modify it and try out another model, a refinement of the old one.

That's what science is all about. Einstein doesn't completely throw out Newtonian physics—it's still in there, but he refines it by making a local perturbation.

Can a machine perform a local perturbation? Not today. But I can envision how it can be done. A machine that decides what experiments to perform next should also be able to modify its theory and continue to progress. That's how I think general AI will eventually become smarter than scientists.

MACKENZIE: Even AI researchers agree we need ethical guidelines for the use of AI. What guidelines would you recommend?

PEARL: I have to answer this question at two different levels. First, at the level of ChatGPT, it's already dangerous because it can be misused by dictators or by greedy businesses to do a lot of harm: combining and distorting data, using it to control a segment of the population. That can be done even today with ChatGPT. Some regulation is needed to make sure the technology doesn't fall to people who will misuse it, even though it's in the very early stage of development. It's not general AI yet, but it still can be harmful.

The second danger is when we really have general AI, machines that are a million times more powerful [than humans]. At this point I raise my hands and say we don't even have the metaphors with which to understand how dangerous it is and what we need to control it.

I used to feel safe about AI. What's the big deal? We take our chances with teenagers, who think much faster than us. Once in a while we make a mistake and we get a Putin, and the world suffers. But most of the time, education works. But with AI, we are talking about something totally different. Your teenagers are now a hundred million times faster than you, and they have access to a hundred million times larger space of knowledge. Never in history has there been such an acceleration of the speed of evolution. For that reason, we should worry about it, and I don't know how to even begin to speak about how to control it.

MACKENZIE: But didn't we talk about this in *The Book of Why?* We discussed the concept of regret, the idea that a machine with a causal model could compare what happened with what would have happened if it took a different course of action. Do you still think regret can equip a machine to make its own ethical judgements?

PEARL: Regret and responsibility will of course be part of AGI and will be implemented eventually using counterfactual logic. Where it will go, I don't know. No matter how well we program the guards of responsibility for this new species, it might decide it wants to dominate the world on its own. It happened to *Homo sapiens*. We extinguished all the other forms of human, the Neanderthal and *Homo erectus*. Imagine what a machine 10 million times smarter could do. It's unbelievable.

The idea of dominating the world could be one of those local perturbations I talked about. The machine might try it out, decide it's fun, and pursue it with vigor.

MACKENZIE: So are you pessimistic now about giving AIs human-compatible ethics fast enough?

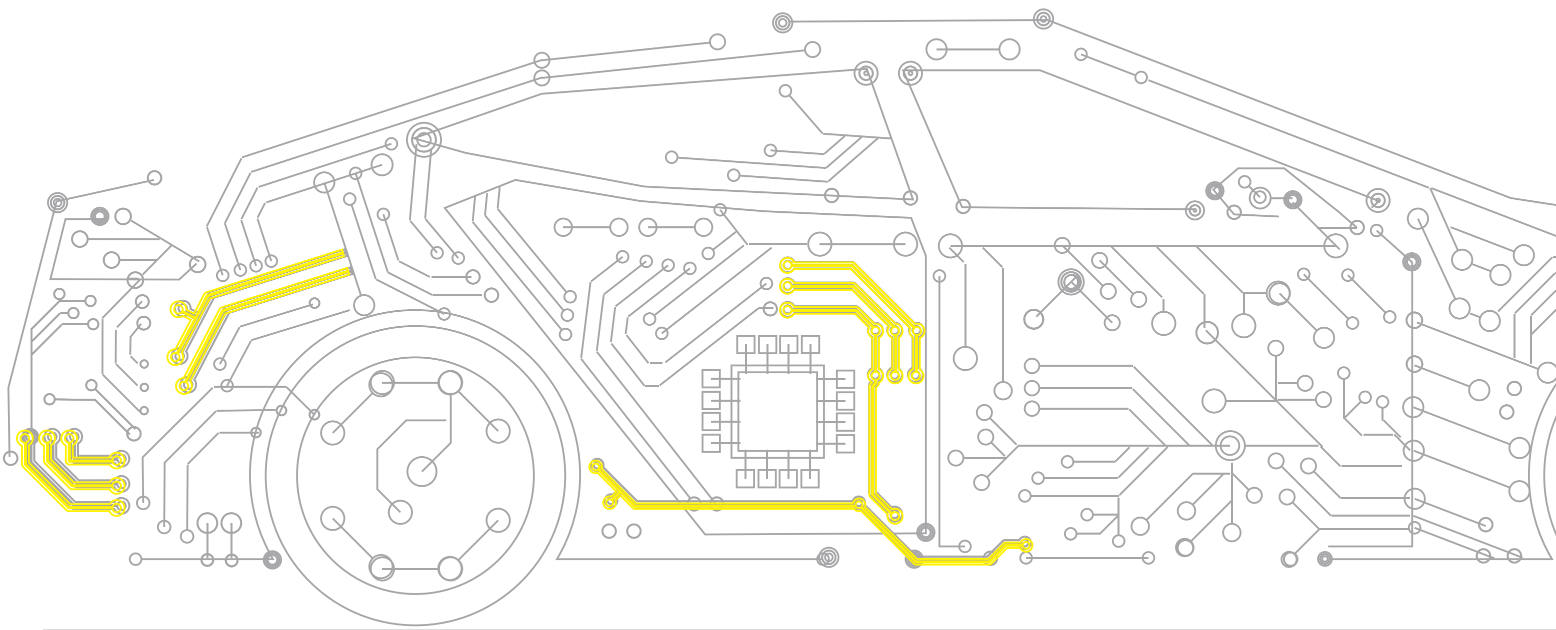
PEARL: You can try to form a committee to regulate it, but I don't know what that committee will do.

MACKENZIE: To conclude the interview, do you have any predictions about what we are going to see in AI in the next year or five years?

PEARL: Do you want to ask me what we are going to see, or what I'd like to see? I'd like to see a shift in emphasis from machine learning to general AI. ChatGPT actually slowed down our progress toward general AI. More and more of our resources will be poured into that direction and not into the correct way of doing AI.

MACKENZIE: But maybe that's a good thing. You said general AI is something to worry about.

PEARL: Here, I am torn. Maybe it's a blessing that ChatGPT is so stupid and society is so intoxicated with it. So maybe we are safe from the danger of creating the new species I mentioned. ■



STATISTICS, AI, and AUTONOMOUS VEHICLES

David Banks and Yen-Chun Liu

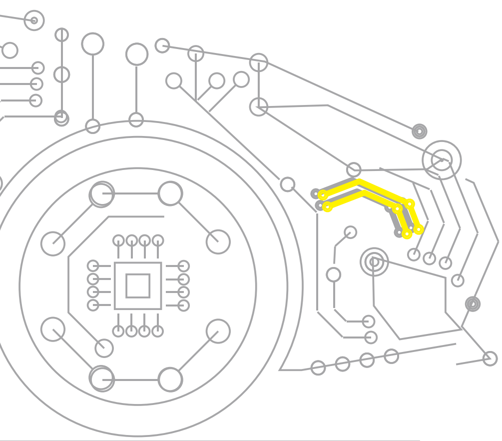
Yogi Berra said, "It's hard to make predictions, especially about the future." But making quantified predictions is something statisticians are supposed to do pretty well.

In this case, we want to examine the future of statistics in artificial intelligence for autonomous vehicles. The exact service statisticians eventually provide will depend on the interplay of technology, law, economics, and social adoption of AVs. We shall be called upon to do risk analyses, probably under many scenarios (e.g., for a mixed fleet and in bad weather). We shall make estimates of the impact of AVs on the environment and economy. We are likely to play a role in insurance and regulation. We may be asked to develop procedures to assess the quality of AI software and the completeness of AI training.

AVs could be transformational. They can significantly reduce pollution, save lives, and provide other benefits. They are one of the few technologies with the potential to meaningfully mitigate climate change.

The greatest benefits would accrue if all vehicles on the road were networked AVs. There would be little need to brake, which is wasteful of fuel—cars could adjust speeds to interweave without stopping and restarting at stoplights. Platooning would also save fuel. And if AVs become sufficiently safe, we no longer need to carry around 1.5 tons of steel for protection. Car bodies could be made of canvas, saving yet more fuel. In terms of safety, AVs are never tired or distracted, and they have better sensors than human-piloted vehicles. If networked, the AVs can share information about safety conditions such as a herd of deer near a road.

According to the National Highway Transportation Safety Administration, there were 42,939 crash fatalities in 2021. The National Safety Council reports that about 20,000 involved multi-car collisions. If only networked AVs were on



the road, this number would presumably drop to nearly zero.

Among the 9,026 single-vehicle fatal crashes in 2021, 59 percent of drivers had blood alcohol levels above 0.08. Not all these fatalities were due to alcohol use, but NHTSA has much evidence to suggest alcohol is a dominant factor in vehicle deaths. Networked AVs would also eliminate deaths and injuries caused by inexperienced teen drivers and elderly drivers with diminished capability.

Modern work using software to pilot robotic vehicles began in 1984, when William "Red" Whittaker launched the NavLab project. In 1995, NavLab 5—a small truck carrying a computer—drove from Pittsburgh to San Diego, mostly without human guidance. Another roboticist involved with that team at Carnegie Mellon University was Sebastian Thrun, who played a leading role in Google's initial effort to build AVs, which has now been spun off as Waymo.

Following are the six levels of vehicle automation:

LEVEL 0. The human has only standard assistance (mirrors, rear-view cameras).

LEVEL 1. Software assistance. Examples include adaptive cruise control and lane keep assist. This level became widely available after 2018.

LEVEL 2. Partial automation. The driver must be hands-on and ready to take control, but the car controls speed and holds its lane. The Tesla Autopilot is an example.

LEVEL 3. Conditional automation. Hands are off the wheel, but the driver must still be ready to control. It is intended for limited access highways and good driving conditions. Many automobile companies are experimenting with this level.

LEVEL 4. High automation. Driver can sleep after inputting destination. Waymo is testing these kind of AVs. The car must stay on traditional roads.

LEVEL 5. Full automation that enables navigation of non-traditional roads.

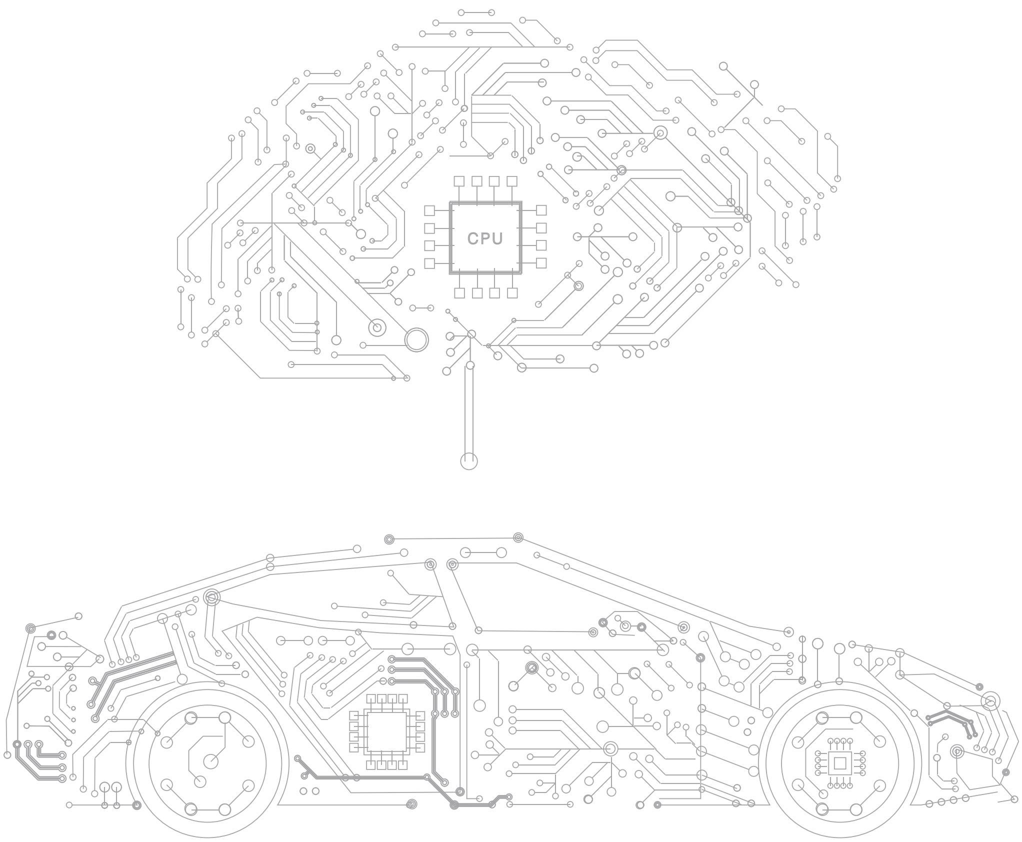
From a statistical risk analysis standpoint, we are beginning to acquire significant data on Levels 3 and 4.

Waymo is essentially the only company operating Level 4 vehicles. They report that, as of January 2023, they had driven a million AV miles. The news release states there have been no injuries and 18 minor contact events. Of those, 10 involved a human driver hitting a stationary Waymo AV. It asserts the human operator violated road rules in every vehicle-to-vehicle collision.

It is difficult to compare the Waymo record to human

David Banks earned his master's in applied mathematics from Virginia Tech, followed by a PhD in statistics. He won an NSF postdoctoral research fellowship in the mathematical sciences, which he took at the University of California, Berkeley. In 1986, he was a visiting assistant lecturer at the University of Cambridge, and then he joined the department of statistics at Carnegie Mellon in 1987. Ten years later, he went to the National Institute of Standards and Technology, then served as chief statistician of the US Department of Transportation, and finally joined the US Food and Drug Administration in 2002. In 2003, he returned to academics at Duke University.

Yen-Chun Liu is a second-year PhD student in statistics at Duke University with research interests in surrogate modeling, Bayesian optimization, causal inference, and reinforcement learning. Before joining Duke, she earned her bachelor's degree in mathematics at National Taiwan University and master's degree in statistics at National Tsing Hua University, Taiwan. Besides studying statistics, she enjoys engaging in outdoor activities and reading. Liu is also passionate about various social issues, including environmental sustainability and social justice.



performance. The number of miles driven is too small to compare fatalities and injuries are difficult to define. Nonetheless, the Bureau of Transportation Statistics reports there were 2,250,000 roadway injuries and people in the US drove a total of 2.9 trillion miles in 2020.

One would have expected 0.78 Waymo injuries if Level 4 AVs are identical to human drivers, but there was none. There is insufficient evidence to conclude Level 4 vehicles are safer than humans, but it strongly suggests Level 4 AVs are not worse. Note that Waymo operates mostly in the metropolitan areas of Phoenix, Arizona, and San Francisco, California, and injury rates tend to be higher in urban areas.

Obviously, statisticians in the BTS and NHTSA are better placed than we are to do risk analyses that control for driving conditions and the kind of accidents that occur. This is clearly an important role for statisticians, and AV safety should be monitored.

Since July of 2021, NHTSA has required AV manufacturers to report crash data. As of January 15, 2023, NHTSA says carmakers have submitted 419 AV crash reports. Of these, 263 have involved Level 2 vehicles, with 156 involving Level 3 or higher AVs. NHTSA further reports 18 fatalities, all with Level 2 cars. No carmaker has reported a fatality with a Level 3 or higher AV, but it should be noted there are far fewer of these on the road. The California DMV reports Level 3 AVs drove a total of 10.4 million miles between January 2021 and December 2022, but that is too small to warrant a risk assessment, since the fatality rate per 100 million miles driven is only 1.34 in 2020.

The NHTSA report has 19 accidents for which the injury level is listed as "unknown." Also, NHTSA does not capture whether the AV was at fault in the crash, nor whether the accident was caused by user error or a problem with the AI. Nonetheless, statisticians at NHTSA are uniquely situated to undertake risk analyses that flag common failure modes.

All current risk analyses pertain to the mixed fleet situation, but the greatest safety benefits accrue when all vehicles on the road are networked AVs.

SCENARIOS

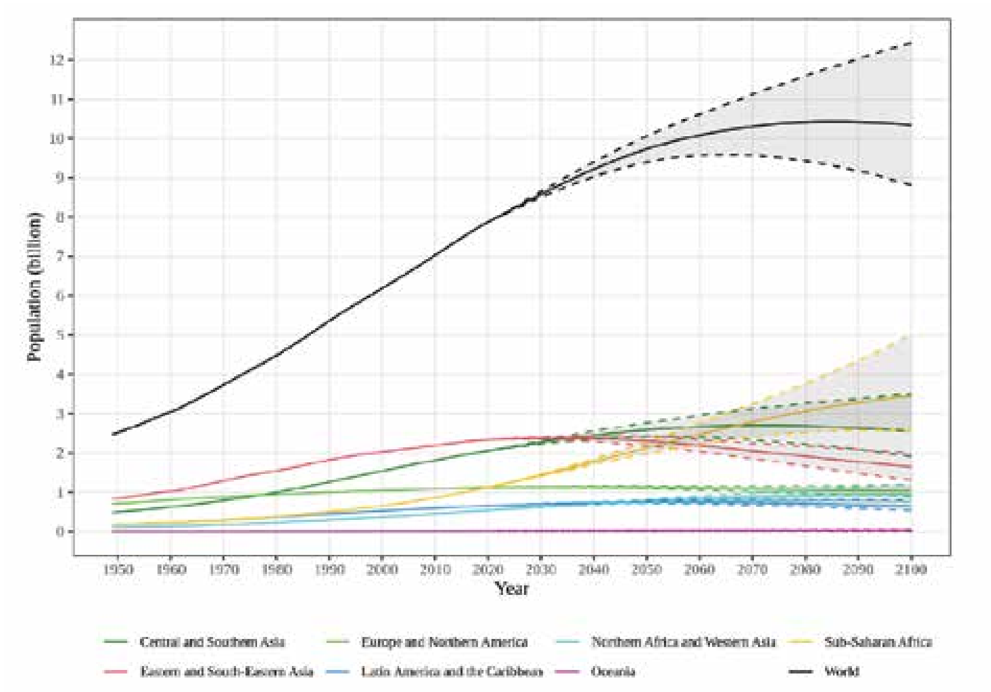
There are several ways AV use might develop. One is the mixed fleet situation we have now. The data hints that, on average, AVs are somewhat safer than humans. This scenario will evolve as car manufacturers and statisticians learn more about how specific levels of AV perform. We can imagine that one day a person might unlock their Level 2 AV and the car will say, "You will have to drive yourself today. There is snow on the road and the AI feels it is not able to operate the automobile safely."

This scenario will lead to new regulations, legal decisions, and insurance policies. The regulations would probably prescribe how software updates should be made securely in a world where international cyberattacks have become common. Similarly, they would govern inspection frequency, the precise capability of different levels of AV, where AVs could be driven, and so forth. Legal decisions would determine liability in crashes, which in turn would drive new forms of insurance policies.

A second scenario, not incompatible with the first, is that a push for AV adoption will come from the trucking and/or ride-share industries. Both will profit if driving can be automated (with additional benefits from all-electric fleets). Perhaps one lane of the interstate highway system will become dedicated to AV trucks. When an AV truck leaves the highway, a human operator might take control, as Air Force pilots direct international drones from domestic bases. Similarly, Waymo is operating a ride-hailing system in Phoenix and San Francisco. Change may happen in tandem; Uber is building a freight fleet that uses Waymo's self-driving trucks.

A third aspect of an AV future is that many vehicles will probably look different than transportation built to move humans. During the pandemic, there was an increase in online purchases, which led to deliveries that encountered the "last mile" problem. One can efficiently move goods from the site of manufacture to the depots near purchasers, but then panel trucks are needed for delivery. These panel trucks might be replaced by AVs that look like an electrically powered grocery store shopping cart.

It is entirely possible that the world will move away from the concept of car ownership. Instead, people might gravitate toward a ride-hailing system. A family of six could summon a large van when traveling together, a fuel-efficient single-seater when only one person is going, or a luxury vehicle when traveling a long distance. Additionally, a ride-hailing system lends itself better to the use of electric vehicles since the AV can take itself off the road to recharge.



Recent and projected human population of the world, broken out by region

TRAINING AN AV

All AVs use deep neural networks. Training a deep neural network requires lots of data, an appropriate architecture, and an optimization algorithm. The architecture is often modular, with subnetworks trained to perform specific recognition or prediction tasks. The optimization algorithms are proprietary but probably quite complex and specific to given subproblems. A limiting factor in training neural networks is they need human beings to categorize unrecognized images such as a heavily weathered stop sign.

There are constraints. Latency must be short so the AI can react quickly to changing traffic conditions. Neural network code must be simple enough to store onboard. The implementations evolve over time and are highly proprietary. Statistical thinking arises in the training of deep neural networks, but it is not central, and so our review of this aspect of AVs is brief.

Initially, the goal was to have an AV that held its lane and followed a car in front at a fixed distance. But now, more complex behavior is wanted. In the Tesla AV, there are multiple camera sensors whose raw images are processed by a rectification layer in the deep network to correct for miscalibrated cameras. These images are sent to a residual network that processes them into features at different channels and scales—features might correspond to stop signs, lane markings, and other vehicles. Features are fused into multiscale information by a module that represents it in a vector space. The output space is

Training an AV AI can be gamed. Applying a Post-it note to a stop sign can fool an AI into thinking the stop sign is a billboard.

sent to a spatio-temporal queue processed by a recurrent neural network. The output is fed to a component called the hydranet, which sends the processed data to modules that handle different prediction tasks. Inez Van Laer's "Tesla's Self-Driving Algorithm Explained" (*bit.ly/3qnRPnn* has more detail, including a link to a YouTube video on the topic by Andrej Karpathy, Tesla's director of AI and autopilot vision.)

Training an AV AI can be gamed. Applying a Post-it note to a stop sign can fool an AI into thinking the stop sign is a billboard.

Regarding training, it is worth noting that Tesla has driven far more miles than Waymo or any other AV system. As of July of 2022, it had driven 35 million miles. In 2019, it had a fleet of around 5,000 vehicles and drove as much in a day as Waymo has driven to date. When a Tesla AI encounters a new situation, or 'sees' something it doesn't recognize, it records it to add to the training database. This means it has more data, thus the potential to jump from Level 3 to Level 4.

IMPACT

AVs will have a major economic impact. They lower the cost of production (e.g., less expensive supply lines, fewer distribution costs). There is more fuel efficiency, and drivers are not paid to pilot trucks. Such change would cause significant dislocation in employment, and an ethical society should plan how to ameliorate the impact on people who lose jobs for the benefit of all.

AVs will have a significant social impact. We may shift from personal cars to a ride-hailing system. Children could be delivered to school by an AV. The elderly could stay in their homes after no longer being able to drive. Before COVID-19, the US Census Bureau estimated the average time spent commuting was 55.2 minutes per day. AVs would allow people to read, work, or sleep, thus giving them an extra hour each day.

But the economic and social impact of AVs is dwarfed by their potential environmental impact. The world's population is increasing. According to the Census Bureau, there are 7.97 billion people now and there will be 9.5–10 billion by 2050.

Researchers at the Australian Academy of Science say the carrying capacity of the planet with anything like a middle-class American lifestyle is on the order of 2 billion. There is no technology on the horizon that can convert a ton of sand into a ton of food and use little energy to do it. But that is the scale of the problems we face.

Climate forecasting is less accurate than demography, but scientists at NASA estimate parts of South Asia, the Persian Gulf, and the Red Sea will be too hot to support human habitation by 2050. And by 2070, the same will be true for parts of Brazil, eastern China, and southeast Asia. Geopolitically, there will be consequences. The people of Bangladesh (who also face problems from the rising sea level) will have to move north to Pakistan or Afghanistan—a relocation fraught with peril.

Similarly, the political governance in Syria and Lebanon is precarious. Climate pressure—say a week of 120-degree temperatures—could lead to collapse and further waves of refugees. The only way for people in such regions to survive is through access to potable water and technology. Many nations in those regions do not have such capacity now and may not acquire the resources in time.

No government in the world has the political will to leave coal in the ground unburned when it is 120 degrees outside and its people need to run a desalinization plant. That sets up a vicious cycle in which more carbon is produced, causing higher temperatures, which leads to more carbon being released.

Which brings us back to AVs. They are one of the few technologies that could have a meaningful impact on our future's carbon footprint. The US Environmental Protection Agency says transportation accounts for 29 percent of US greenhouse gas emissions; it is the largest single component of our pollution budget. In the United States, transportation accounts for 41.7 percent of CO_2 emissions (77 percent if one excludes wildfires). Networked, electric AVs could dramatically mitigate global warming while simultaneously improving our quality of life socially and economically.

AVs have enormous potential impact, and forward planning can achieve important social and economic benefits, as well as essential environmental benefits. AVs depend upon AIs, which are trained upon vast quantities of data. Statistical methods are key to many aspects of AV implementation. Risk analysis is obvious, but our community can and should play a much larger role. ■

MORE ONLINE

Find an academic version of this paper, complete with references showing heavy use of work by statisticians from several government agencies, at *www2.stat.duke.edu/~banks.*

AI–ENABLED MEDICAL DEVICES and DIAGNOSTICS: STATISTICAL CHALLENGES and OPPORTUNITIES FROM A REGULATORY PERSPECTIVE

Gene Pennello and Frank Samuelson

Artificial intelligence is poised to deliver important contributions to medical device application areas, including image acquisition and processing; earlier disease detection; more accurate diagnosis, prognosis, and risk assessment; identification of new observations or patterns of human physiology; development of personalized diagnostics and therapeutics; and treatment response monitoring, to name a few. However, the complexity of medical device AI algorithms, the data-driven nature in which they are trained, their rapid application to many medical areas, and the unique nature of clinical medical data (e.g., low prevalence of disease, lack of or difficulty in obtaining truth data, etc.) create challenges in developing robust evaluation methods for AI devices. These include clinical and nonclinical testing and understanding the impact of these devices in the real world. In addition, some medical devices may employ AI algorithms designed to learn as data accumulates, which presents unique evaluation challenges.

According to the US Food and Drug Administration public listing, at least 522 AI-enabled medical devices have been marketed in the United States as of October 5, 2022. This data and a 2020 review by Stan Benjamens, Pranavsingh Dhunnoo, and Bertalan Meskó in their *NPJ Digital Medicine* article, "The State of Artificial Intelligence–Based FDA-Approved Medical Devices and Algorithms: An Online Database," indicate marketed AI-based medical devices are predominately in the field of radiology, followed by cardiology and internal medicine/general practice.

In pathology, the first marketed AI device is Paige Prostate, a software device applied to digital histopathology images of prostate needle biopsies that uses a neural network to classify an image as suspicious or not for prostate cancer. When suspicious, Paige Prostate provides a single coordinate (X,Y) of the location with the highest probability of cancer for further review by a pathologist.

In radiology, medical devices incorporating AI have expanded from assisting radiologists in segmentation or detection (CADe) to quantitative imaging, computer-assisted diagnosis (CADx), triage, and multi-class classification. One example of a quantitative imaging device is the caption interpretation automated ejection fraction software, which applies machine learning algorithms to process echocardiography images and estimates left ventricular ejection fraction.

QuantX is an early example of an AI CADx device that assists in the characterization and diagnosis of breast abnormalities. The device automatically registers, segments, and analyzes user-selected regions of interest in magnetic resonance images of the breast to extract morphological


Gene Pennello is a statistician and **Frank Samuelson** a physicist for the US Food and Drug Administration Division of Imaging, Diagnostics, and Software Reliability. The division conducts research on methods for evaluating performance of medical devices, including medical imaging systems, diagnostic tests, and medical devices enabled with artificial intelligence / machine learning algorithms. Technical disciplines of the staff include physics, electrical and biomedical engineering, mathematics and statistics, computer science, and medical radiology.

features (e.g., lesion area, sphericity, homogeneity, volume, contrast, etc.) and radiomic features. These are then analyzed by an AI algorithm to obtain a QI score, which confers relative likelihood of malignancy.

Other examples of CADx devices that combine regions of interest detection functions include OsteoDetect and FractureDetect, which use ML to analyze adult radiographs of various anatomic areas to identify and highlight potential fractures while providing additional diagnostic information.

Computer-assisted triage and notification (CADt) devices create an active notification to providers for cases identified as likely containing a time-sensitive finding, giving them the option to move such cases to the top of the reading queue. Cases unflagged by CADt devices are read without priority according to standard of care. By aiding radiologists in identifying cases that should be given high read priority, CADt devices may provide benefit when early detection of the target condition is crucial for effective intervention.

Three examples of AI CADt devices are ContaCT, BriefCase, and Viz ICH, which are applied to computed tomography angiograms of the brain or head. ContaCT notifies neurovascular specialists of a potential large vessel occlusion stroke, while BriefCase notifies hospital networks and trained radiologists of potential large vessel occlusion stroke. Viz ICH notifies hospital networks and trained clinicians of potential intracranial hemorrhage stroke.

CADt devices are evaluated not just for accuracy in detecting the target condition but also for time saved in detecting the target condition earlier in cases from patients most likely to benefit from an earlier image interpretation. Yee Lam Elim Thompson, Gary Levine, Weijie Chen, Berkman Sahiner, Qin Li, Nicholas Petrick, Jana Delfino, Miguel Lago, Qian Cao, Qin Li, and Frank Samuelson applied queueing theory to develop estimators of the mean time saved by CADt devices among cases with the target condition in "Evaluation of Wait Time Saving Effectiveness of Triage Algorithms."

Most marketed classification devices distinguish between two states of health: presence or absence of a target condition. Multi-class classification devices distinguish between more than two states of health. qER is a CADt device that applies classical ML and a deep convolutional neural network to voxels on brain CT to detect intracranial hemorrhage, mass effect, midline shift, and/or cranial fracture. Additionally, assays are being developed to screen for multiple cancers by detecting circulating tumor DNA in plasma samples. When a cancer is detected, ML is employed to classify its origin and focus appropriate diagnostic work-up.

Unfortunately, many developers have not yet taken full advantage of ML for multi-class classification, opting instead to train binary classifiers for each condition separately and bundle them into a single device. Challenges with evaluating multi-class classification devices include designing an efficient study for evaluating device clinical accuracy for multiple conditions—especially when some have low prevalence—evaluating the benefit-to-harm trade-offs of true and false test positives and true and false test negatives for each condition, and developing an appropriate statistical analysis plan for multiple hypothesis testing of each condition and possible combination of conditions.

AI is also being used to develop medical devices that quantify the risk of developing a target condition by a future time. Risk prediction models are evaluated for how good the predictions are. Calibration refers to how well the number of events predicted agree with the number of events observed in a prospectively sampled cohort. Risk predictions should be well calibrated, otherwise they may lead to inappropriate clinical management. In "Calibration of Prognostic Risk Scores," published in *Wiley StatsRef,* Ben Van Calster and Ewout Steyerberg define mean, weak, moderate, and strong risk calibration. An open question is which definition of risk calibration should be considered when developing acceptance criteria for validating risk predictions as well calibrated.

Currently, most devices employ fixed algorithms that provide the same output each time the same input is provided. Soon, however, AI devices—particularly those employing ML—may be designed to be updated periodically or continuously as they learn from accumulating data.

How to evaluate the performance of a learning medical device is an open question. The FDA Center for Devices and Radiological Health issued the 2019 discussion paper, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)–Based Software as a Medical Device," in which CDRH describes a potential approach for how updates to AI/ML software as a medical device, including those based on periodic or continuous learning, could be regulated. CDRH envisions using a *predetermined change control plan* that pre-specifies the scope of the anticipated device modifications and an *algorithm change protocol* specifying procedures and testing to be followed after each modification is made so the device remains safe and effective.

Post-market monitoring may be important for mitigating the risk of AI device performance deterioration. For example, performance may deteriorate as data drifts for devices with fixed ML algorithms, which may occur as populations or medical practices change over time. For continuously learning devices, an unrepresentative data stream could bias the learning process, with concomitant performance deterioration in device updates.

Jean Feng, Scott Emerson, and Noah Simon proposed statistical procedures for monitoring a sequence of device updates in "Approval Policies for Modifications to Machine Learning–Based Software as a Medical Device: A Study of Bio-Creep," published in *Biometrics.* In "Monitoring Machine Learning (ML)–Based Risk Prediction Algorithms in the Presence of Confounding Medical Interventions," Feng and Alexej Gossmann, Gene Pennello, Nicholas Petrick, Berkman Sahiner, and Romain Pirracchio proposed procedures to monitor devices for changepoints in real-world clinical accuracy and utility after adjusting for confounding medical interventions.

While the potential for performance deterioration is possible with any medical device, it is worth special consideration for AI devices given the data-driven nature in which they are trained.

Many issues with training and validating AI models for health care have been discussed in the literature. Some of these issues include the following:

1. **Uninterpretability** of "black box" ML models, especially for making high-stakes decisions
2. **High uncertainty** of predictions from unregularized models fitted in a high-dimensional prediction space
3. **Reproducibility** of AI model results (i.e., variation across repeated measures of the input data)
4. **Confounding,** an AI model's reliance on a spurious association with outcome
5. **Nonrepresentative training data** leading to bias, poor overall out-of-sample performance, and/or lack of **generalizability** of performance across subgroups (i.e., fairness), the lack of which may lead to health care disparities
6. **Nonstandard data** for training or validation:
    - **Synthetic data** generated by a generative adversarial network or other generative model trained on private data to enable data sharing with differential privacy guarantee
    - **Data reuse,** validating an updated model on the same data set on which the original model was validated
7. **Imperfect reference standard**
    - **Misclassification** of ground truth of the target condition in some subjects (e.g., **natural language processing** may be imperfect for deriving phenotype ground truth)
    - **Weak supervised learning** from training data with some reference ground truth labels missing or incorrect or with coarsened reference ground truth labels

8. **Performance deterioration** of a fixed AI model because of **data drift,** for example, or of a learning AI model because of highly unrepresentative or **adversarial cases** appearing in the data stream, for example

Many medical device AI algorithms have so many parameters that they are essentially uninterpretable, hence sometimes called "black box." For example, in radiology, a deep neural network employed for a medical imaging task may involve a huge number of parameters consisting of weights and biases embedded in multiple hidden layers. Evaluating whether an uninterpretable AI algorithm provides clinically significant results in validation data may be problematic because clinical understanding of the parameter space, much less the algorithm itself, is lacking. In particular, the more complex an algorithm is, the more likely it has complex failure modes that are hard to identify, especially in smaller size studies commonly used for validation.

In contrast, an assay for a target condition based on measuring a single biomarker in a well-understood specimen matrix (material or medium within which the biomarker is measured such as serum, tissue, or fluid) using stable *in vitro* diagnostic products (e.g., reagents, calibrators, and controls) can be relatively easy to understand clinically. The single biomarker simplifies the evaluation of whether the assay provides clinically significant results, increasing the likelihood it will be adopted into clinical practice.

Attempts have been made to explain uninterpretable AI output by using perturbation or gradient-based methods post hoc to obtain metrics (e.g., LIME, SHAP, or GradCAM) that quantify which input variables contributed the most to the output.

However, Satyapriya Krishna, Tessa Han, Alex Gu, Javin Pombra, Shahin Jabbari, Steven Wu, and Himabindu Lakkaraju indicate in "The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective," current methods often disagree about which input variables were most important to an output. Whether disagreement of explainability methods indicate deficiencies in these methods or lack of robustness of the particular AI model appears to be an open question. Similarly, Amirata Ghorbani, Abubakar Abid, and James Zou, in "Interpretation of Neural Networks Is Fragile," published in the *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence,* showed feature importance maps that explain which features in an image led a neural network to a particular classification can be unstable.

Absent clinical understanding of a complex AI device algorithm, statistics, then, may need to provide the foundation for performance evaluation. Fortunately, many of the same statistical principles employed for the design, conduct, and analysis of performance evaluation studies of medical products in general may be applied to AI devices in particular.

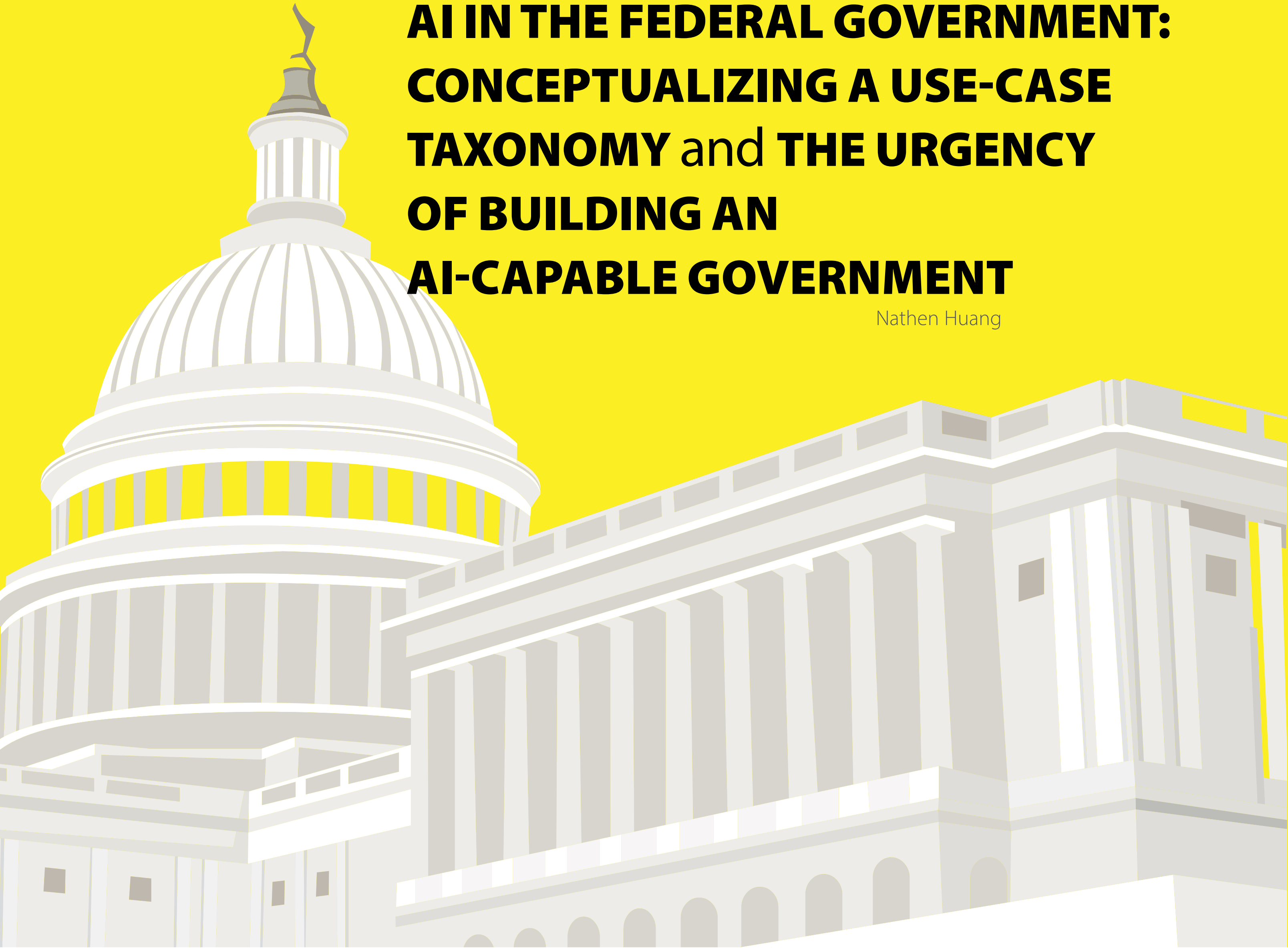
AI algorithms will play an increasingly important role in medical devices for the foreseeable future. However, enthusiasm for medical device AI could be tempered if challenges are not addressed with development (e.g., high uncertainty of some AI outputs) and evaluation (e.g., of learning devices). Clinical uses of AI in medical devices are anticipated to grow and may only be limited by the imagination of developers and data available. With new clinical uses will come new challenges in design, conduct, and analysis of performance evaluation studies, creating opportunities for statisticians to develop novel designs and evaluation methodologies to address each new clinical use. Many opportunities exist for statisticians to play a role in AI algorithm development and evaluation. ■

MORE ONLINE

For a complete list of references and additional reading material, visit *Amstat News* online at *https://magazine.amstat.org.*

AI IN THE FEDERAL GOVERNMENT: CONCEPTUALIZING A USE-CASE TAXONOMY and THE URGENCY OF BUILDING AN AI-CAPABLE GOVERNMENT

Nathen Huang



When I was in college, I grew entranced with a somewhat propagandistic show called *Parks and Recreation*. The show follows Leslie Knope, an eager and overly dedicated manager of a team of quirky bureaucrats who manage the daily operations of running the parks department in the insignificant suburb of Pawnee, Indiana. While the show primarily showcases and satirizes the hijinks and inefficiencies of municipal government, I came away with a different impression: Those working in the public sector had admirable and unbreakable resolve to provide better services to people, regardless of their background.

In the years since I first watched *Parks and Rec*, I've found myself drawn to public-sector work and have worked for the federal government as both a contractor and civil servant across various agencies: Department of Transportation; Department of Homeland Security; Department of Defense; Centers for Medicare and Medicaid Services; Department of Housing and Urban Development; and (now) Department of State.

Having worked with data in various federal capacities my entire career as a data scientist, I have seen firsthand both opportunities and challenges for greater data science capacities—as well as the subsequent popularization of artificial intelligence—in the federal government. I have developed a taxonomy that summarizes AI's usefulness to the government for three primary use cases.

CASE MANAGEMENT

First, the most direct application of AI in the federal government is in case management. I use the term "case" loosely; a case is simply any instance of a file, application, document, issue, or matter routinely dealt with in the government. As the federal government is meant to be a responsible steward of citizens' data, cases are naturally the biggest challenge the government faces—and its most plentiful source of data. Deciding which civil rights complaints are prosecutable; determining eligibility for subsidies in housing, veteran health care, or taxes; and identifying which refugees are immediately eligible for asylum vs. parole are all instances of cases to manage.

Much of the government does this manually. Not only is it tedious and laborious, but reviewing individual cases according to a set of standards often open to interpretation can introduce all sorts of bias—availability bias, racial bias, selection bias, etc. Nonetheless, case management will always be essential to governmental operations if citizens need their individual situations brought before the government and addressed.

AI has a lot to offer here. By automating and making predictions on data, case adjudication can be much easier and judged less inconsistently. However, that doesn't mean AI isn't immune to faulty decision-making—after all, machine learning algorithms are only as good as the data they are trained on—and we should be vigilant to AI's harms. While using AI systems for judging cases, the government should prioritize data privacy and security in algorithms given the risks inherent in revealing and using citizen data—especially if training data can be accessed or exposed. If we are thoughtful and careful about how we deploy AI systems in reviewing cases—mitigating bias as much as we can—we will benefit from algorithms that make case management much more consistent and accurate.

EFFICIENCY IN DELIVERING SERVICES

Another benefit of AI is it can help the government deliver more efficient services. Beyond simply being able to manage various cases that come through the government, AI systems can help the government push out critical services to its citizens more quickly. Problems the government works on in which speedy services are necessary include serving populations at greatest risk for certain diseases and in greatest need of medication and treatment; anticipating which students and businesses should be eligible for loans and tax benefits; predicting and responding to disasters with equipment and adequate staff—for both health-related crises like COVID and natural disasters like earthquakes and hurricanes—and identifying who will be affected.

Nathen Huang (he/him) is a quantitative social scientist and data scientist who is passionate about leveraging statistics to advance society's well-being. After years of working in federal consulting, Huang was detailed to the US Department of State as a Presidential Innovation Fellow and subsequently joined the federal service as a program analyst. He is a graduate of Columbia University's MA in quantitative methods in the social sciences program and lives in Washington, DC.

When executing decisions manually, government services can operate slowly and inconsistently; for each case, there may be myriad ways of activating relevant systems to get assistance to the people who need it. At the very least, AI can offer a programmatic and potentially less-biased system for matching the right federal programs to the people who need them, removing the bureaucratic red tape and intermediaries required to make a decision.

AI's potential for efficiency should inspire us to more quickly identify ways to incorporate AI into federal government practices, but it should also give us pause. When the impact of AI systems is on the entire country, we must consider the risks and dangers of

THE FLIP SIDE

When we approach algorithmic bias, our instinct is often to ask how we can create less biased algorithms—after all, biased algorithms can perpetuate the existing biases we hope AI will resolve in a more impartial manner. However, designing less biased algorithms is just one aspect of addressing the larger challenge of AI bias, especially when it comes to federal work that affects the entire country. We should ask not just how to create less biased algorithms, but how biased algorithms create the conditions for future bias.

For instance, algorithms trained on specific ethnic groups, binary gender, or geographically unrepresentative areas could be used to create recommendations that reinforce inequalities in our society. A job recommendation algorithm lacking training data on applicants with a disability or graduates of historically Black colleges and universities could cultivate a less diverse workforce, perpetuating the false belief that able-bodied people and/or persons of color are not qualified workers.

Moreover, rather than making the federal government better, AI in the government may instead cultivate limitations on how we view humanity—in binary or limited categorical outcomes. In this way, we may become dependent on AI to moderate our understanding of truth, rather than conceptualizing the world in a way that aligns with how we, as human beings, live and behave. As the saying goes, "Computers are binary; people are not."

AI systems are great for helping the federal government, but they are not the solution to all the challenges our government faces. Because we know algorithmic bias can emerge from so many places—pre-existing federal standards, data sources, political and legislative pressures, IT systems—we should consider how we can have a role in shaping the future we seek to create.

'getting it wrong'—especially if AI causes federal programs to move too quickly, offer shoddy services, or engage in activities that can harm privacy or exceed legal boundaries.

PROCESS OPTIMIZATION

Finally, the third benefit to AI in the federal government is the opportunity for process optimization. Unlike service delivery efficiency, process optimization focuses on ensuring the government achieves a goal in an "optimal" way—which may or may not ensure overall "efficiency" of the service but focuses on doing a task better. For example, weapons and vehicles fitted with AI-powered systems optimized for various terrains and conditions enable the sub-agencies of the Department of Defense to conduct missions remotely and quickly respond to threats. Likewise, with the looming deployment of autonomous vehicles, there is a critical need for the government to ensure these vehicles are cooperating with each other on the road and following federally set standards.

AI can also be useful for detecting threats to the country. For the FBI and intelligence agencies, AI systems can quickly detect when fugitives are most likely to commit crimes and enter or flee the country based on online activity and communications gathered from signals intelligence. While many of these processes are not easily optimized—since optimization requires a standard that can be validated for whether an objective was achieved—they offer examples of how AI learning could improve processes in automatic ways that enable the government to respond more quickly to and improve upon the various problems it addresses.

Unfortunately, while harnessing AI systems appropriately poses a challenge, one of the greatest barriers to the federal government leveraging their potential is the government's inability to recruit and retain necessary talent to deploy them, which is somewhat inherent in the nature of the work. When agencies, bound to previous standard operating procedures, are unable to conceptualize how advanced data analytics capabilities can benefit existing federal workstreams, talented people may not be motivated to find jobs in the government.

Compensation is also an issue. When technology companies are paying 100–200 percent of the entry-level federal salaries for engineers and data scientists, even the most well-meaning technologists may not be motivated enough to work for the government.

To incentivize technologists to invest their talents in the public good, the Presidential Innovation Fellows program has matched private sector talent with high-impact technology projects in the federal government for more than a decade. Increasingly, other programs have emerged to incentivize technologists as well, including the US Digital Corps, Congressional Innovation Fellows, and AAAS Fellowship.

Building an AI-capable government—one that affords greater benefits to society than the government we have now—will take all types of people and skills. The federal government needs AI-competent experts in many areas—including quality assurance engineers, data architects, data engineers, and policy analysts—to reconcile government processes with data services that comprise AI systems. Data stewards and policy experts will be needed to check data intake; maintain processes for funneling data; and ensure the reporting, visualizing, and analyzing of collected data.

Parks and Rec shows the work of helping the public can sometimes be tedious and draining, but it also demonstrates how public service can be immensely rewarding and impactful. When implemented properly and thoughtfully, AI systems can help take away the tedium of public sector work and free up civil servants to exercise their creativity and solve novel problems.

The prospect of doing the most good through public service enabled by technology continues to motivate my work and desire to stay in government. I hope more people feel empowered by that possibility, too. ■

EDITOR'S NOTE

The views expressed in this article are those of the author and do not necessarily reflect those of the US government.

Celebrate Women in Statistics and Data Science This October

The Caucus for Women in Statistics will host the International Day of Women in Statistics and Data Science October 10 to celebrate female statisticians and data scientists around the world.

The aims of the virtual conference are the following:

- To showcase women and their contributions to the field
- To connect women statisticians and data scientists around the world
- To encourage collaborations among statistical societies around the world
- To prompt statistics and data science to become more inclusive and diverse
- To bridge statistics and data science

There will be both live and recorded presentations. Submit session ideas to *idwsds@cwstat.org.com* and follow updates on Twitter by searching for @cwstat.

TOP 10 THINGS I HAVE LEARNED FROM CHATGPT

Michael Hansen

In addition to teaching AP Statistics for 25 years, I have taught high-school computer science for about 13 years. This summer, I used ChatGPT to help me write a Python application of about 3,000 lines. Here are my top 10 takeaways from that experience.

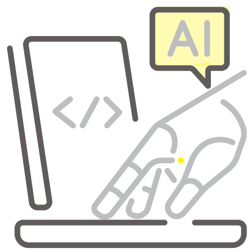
Michael Hansen studied mathematics and computer science at Bradley University (BA) and the University of Illinois at Urbana-Champaign (MS). After working 12 years in the computer field as a government contractor, he joined the faculty of St. Albans School, an independent school for boys in Washington, DC. He plays violin, viola, and piano semi-professionally.

1

ARTIFICIAL INTELLIGENCE IS A GAME-CHANGER FOR SOFTWARE DEVELOPMENT. I cannot imagine writing code in the future without using a chatbot or similar AI tool. It would be like writing a long document without a spell-checker or navigating a car without GPS. Possible? Yes, but not what most modern people would do.

2

WE NEED A COMPLETE RE-THINK OF HOW TO TEACH COMPUTER SCIENCE, or at least the coding aspect of computer science. Coding is no longer the challenge students thought it was. Indeed, ChatGPT writes beautiful code for all but one of the exercises in the Java textbook I use for my beginning students. Paste the text of the exercise into the ChatGPT query window, press enter, and voila! In most cases, you receive code an A+ student would envy. Not only that, but ChatGPT's code includes coherent comments to explain what it is doing. (Try getting your beginning students to do that!)



3

THERE IS NO LONGER A BARRIER TO CODING FOR ANYONE. The reason students formerly thought coding was challenging is it seemed arcane. Now that coding is easy, it hardly qualifies as arcane. In the past, some people thought coding was key, meaning every high-school student should learn how to code. I never thought that, but I did not contradict people who did because they supported a cause near to my heart: computer science education for all. In light of the new AI chatbots that will revolutionize how code is produced, I would say computer science is key, and understanding how to formulate requirements is key, but coding, per se, is not. It is akin to when knowing what the standard deviation measures is key, but knowing how to compute s.d. by hand is not. I still make my students compute s.d. manually a few times in hopes they will see how larger errors contribute more to the statistic, but it is hard to make a case that s.d. computation is a key skill for statistics.

4

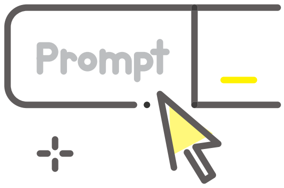
SOME 'OLD SCHOOL' ASSESSMENTS MAY REMAIN VALID, BUT DEFINITELY NOT AS TAKE-HOME EXERCISES. One assessment type I use to gauge student fluency with terminology and notation—namely, fill-in-the-blank—is tailor-made for ChatGPT, which was built from the outset as a conditional probability engine for finding the most likely word or phrase to finish a thought. On a fill-in-the-blank assessment, ChatGPT easily outscores even my best students. Nevertheless, I believe the skill being measured is important for humans to learn, even if AI bots can run rings around them. Tellingly, ChatGPT is already 'better' than many students at scoring well on the measures they care most about, namely the College Board's Advanced Placement exams … and the technology will keep improving each month.

5

IF YOU ARE NOT HAPPY WITH CHATGPT'S INITIAL RESPONSE, ASK IT TO TRY AGAIN. On the second or third try, it almost always does better. However, be aware that ChatGPT always responds with a confident tone, even when it is wrong. On my summer project, I had to reformulate one set of function requirements more than a dozen times, but it was still worth it, since the code ChatGPT produced for me would have taken me hours to research, write, and debug on my own.

6

THE KEY CHALLENGES FOR STUDENTS REMAIN THE SAME: formulating the requirements of a problem in sufficient detail and clarity; testing/recognizing problems in the code (whether written by humans, AI, or a combination); understanding what the computer is actually doing with the code; using that knowledge to debug and refine software; and designing an entire system to begin with. AI helps with the debugging, but the other challenges remain uniquely human—at least for now. I am old enough to remember sales pitches 40 years ago for software products that would supposedly make developers obsolete. What happened instead is the field evolved to use different tools, but not fundamentally different skills. Someth.ing hard always remains; there is no silver bullet for software development. The classic book on the subject, *The Mythical Man-Month* by the late Frederick Brooks, is still in print almost half a century after its release.



7

DEBUGGING IS EASIER. For the most common student bugs (e.g., fencepost errors, lines inside a loop that should be outside the loop or vice versa, improperly copied-and-pasted code that should have been encapsulated, etc.), ChatGPT will usually find the bugs if you paste your code into the query window, even if you do not describe the symptoms. There is grave danger here, because I believe the most educational experience in computer science is the adventure (cough, cough) of spending hours tracking down logical errors in a program and learning—through the school of hard knocks—how to write better code. I fear students, if deprived of the education serious debugging entails, will never fully mature as computer scientists. Then again, I may be as misguided as people in the 1970s who thought students would lose something important by not extracting square roots with pencil and paper.

8

CHATGPT WILL REVEAL TO EVERYONE WHAT COMPUTER SCIENTISTS HAVE KNOWN FOR DECADES: translating requirements from English into computer code was never the central problem. (Proof: Translating detailed requirements from English into code was a task historically given to junior programmers. Now, with the advent of ChatGPT, that task will become largely a matter of copy-and-paste.) The hardest task of all is requirements analysis. Once the requirements of a problem are well understood and formulated at a sufficient level of detail, the problem is mostly solved. For large systems, architecture and communication among the developers (as Brooks noted) become daunting challenges, as well.

9

EVEN THOUGH CHATGPT WILL WRITE CODE, I NEED TO PERSUADE (AND INSIST) MY STUDENTS LEARN HOW TO WRITE THEIR OWN CODE. A certain set of core knowledge is essential in any field. Yes, you could skip learning "Happy Birthday," but you would look rather pathetic if you had to look up the words and music on your phone every time your coworkers gathered around a cake in the company break room. In computer science, the best way to formulate detailed requirements is to write pseudocode, and it is hard to write good pseudocode without having written real code.

10

I NEED TO WRITE NEW GRADING RUBRICS THAT RECOGNIZE THAT MUCH OR ALL OF THE CODE IN STUDENT PROJECTS MAY HAVE BEEN WRITTEN BY CHATGPT, regardless of what I tell the students to do and regardless of any honor-related penalties I may threaten. My time-honored trick of writing project assignments with requirements too specific to permit plagiarism from the internet will no longer work, since ChatGPT has no trouble writing code to satisfy virtually all student-level project requirements. A colleague proposed that for take-home projects, I could have students write detailed requirements and submit those to me without code. The grade would then be based on the quality of the code ChatGPT writes in response and the quality of the enhancements the students suggest (and justify). That idea holds promise because it recognizes the reality of AI in the workplace of the future and leverages that reality as part of an authentic assessment. ■

A STUDENT'S VIEW

EMBRACING AI-DRIVEN EDUCATION: CHATGPT AS A LEARNING TOOL

Jaya Kolluri

In today's rapidly changing world, the integration of AI-based technologies like ChatGPT into education has sparked an ongoing debate between teachers and students. Many educators worry using AI-driven tools for learning may lead to laziness in students and discourage them from conducting their own research. However, I—as a high-school student—think AI-driven models can complement traditional teaching methods. ChatGPT offers students additional resources, perspectives, and the ability to explore different concepts independently. As students, we thrive on curiosity and the desire to discover, and ChatGPT can support and nurture that inquisitive nature.

My experience with ChatGPT convinced me of its true potential. I decided to create an "Introduction to Statistics" blog powered solely by ChatGPT. To my amazement, this AI tool assisted me in drafting engaging, informative, and well-structured blog posts that covered various statistical concepts clearly and concisely. One of the greatest benefits to me was its ability to generate content in minutes, saving me hours of research and writing. Acting as an on-demand teaching assistant, it answered my questions and provided examples that made the concepts more accessible to me. Moreover, it allowed me to present different angles so I could provide a well-rounded post. This highlights AI's role as a collaborator of—not a replacement for—human creativity.

My goal in creating this AI-powered blog was to showcase ChatGPT's potential to be a valuable educational tool when used strategically and urge teachers to embrace it to enhance their teaching methods.

Jaya Kolluri is a rising high-school senior at Winsor School in Boston, Massachusetts. Her research in STEM subjects and data science has led to several conference presentations and journal publications. She is interested in pursuing higher education in health care and applied data science and can be reached at *jaya@kolluri.com*.

Embracing technology in education does not mean abandoning traditional methods. Finding a synergistic relationship between the two can empower both students and teachers to learn more effectively. By fostering a positive outlook on using ChatGPT in education, we can pave the way for a more dynamic and enriching learning experience.

I invite you to explore and interact with my blog at *https://statlog.blog*. Also, share your opinions about the blog's utility as an educational resource in the comments section of each post. ■

ETHICAL CONSIDERATIONS FOR DEVELOPING and USING ARTIFICIAL INTELLIGENCE IN STATISTICAL PRACTICE

Members of the ASA Committee on Professional Ethics

There is much discussion of artificial intelligence in the popular and scientific press these days. One description of AI is that it refers to systems displaying intelligent behavior by analyzing their environment and taking actions—with some degree of autonomy—to achieve specific goals. Most AI relies on data and data-driven algorithms to determine the actions that best achieve the goals.

The ASA Ethical Guidelines for Statistical Practice (*https://bit.ly/452gmgM*) define "statistical practice" as including activities such as designing the collection of, summarizing, processing, analyzing, interpreting, or presenting data and model or algorithm development and deployment. In this way, AI often fits squarely into the space of statistical practice. Whether tuning models, building novel techniques or algorithms to accomplish a task, or mining data for patterns to influence a product life cycle, developing and using AI nearly always includes elements of statistical practice. Naturally, many statistical professionals may have questions about how AI affects their practice.

UNDERSTAND POTENTIAL RISKS AND AVOID HARM

AI's powerful, transformative force and profound effects on every corner of the world are undeniable. Yet we also understand AI is prone to bias, inaccuracy, and unfair outputs. The ASA Ethical Guidelines offer actionable guidance for reducing potential risks of using AI in statistical practice, including seeking to understand and mitigate known or suspected limitations, defects, or biases in the data or methods (see Principle B).

AI can collect, process, and analyze large amounts of personal and sensitive data, which can pose risk to individuals' privacy, autonomy, and dignity. Ethical statistical practitioners should try to reduce harm by respecting privacy, equality, and autonomy of individuals and groups.

Principle D in the guidelines promotes protecting and respecting the rights and interests of people about whom data is collected and those who will be directly affected by their practice. Specifically, it includes protecting people's privacy and confidentiality of data; obtaining proper consent for data use from respondents and data sources when applicable; considering impacts of statistical practice on society, groups, and individuals, particularly those who are disadvantaged; and minimizing adverse impacts from applications or in the reporting of results.

PROMOTE TRANSPARENCY

Most existing AI algorithms have a "black box" nature, where the decision-making process is so complex it cannot be explained in a way that can be easily understood. Despite this, ethical statistical practitioners should seek to adopt plain language to describe the utility, methods, and risks associated with AI. In statistical practice, transparency can include listing data sources, addressing privacy concerns, explaining which methods are used and how they are implemented, and being open about the risks and biases.

For example, Principle B promotes transparency on assumptions made in the execution and interpretation of statistical practice, including methods used, limitations, possible sources of error, and algorithmic biases. Principle C encourages ethical statistical practitioners to inform stakeholders of the potential limitations on use and re-use of statistical practices in different contexts.

When developing an AI algorithm or advocating for the algorithm's usage, ethical statistical practitioners should educate nonstatistical practitioners about how to understand and interpret these limitations and concerns as part of this transparency. Promoting transparency can facilitate trust from collaborators, stakeholders, fellow statistical practitioners, and the public users of statistical products. In addition, it helps minimize harm and improve the utility of AI in statistical practice.

APPLY AND MAINTAIN PROFESSIONAL COMPETENCE

As AI becomes more powerful and takes on tasks with fewer human instructions, it can influence, replace, or override human decision-making and actions, which raises questions about the role and responsibilities of humans in the AI era. Ethical statistical practitioners should apply and maintain their professional competence and keep up to date with the latest developments in AI and statistical practice. Specifically, professional competence involves having the ability and skill to oversee the work, identify possible sources of error and bias, support work with robust methods, and have the domain knowledge to understand questions to be addressed.

Principle A in the ASA Ethical Guidelines calls on statistical practitioners to evaluate potential tasks, assess whether they have (or can attain) sufficient competence to execute each task, and acquire and maintain competence through building skills to maintain a high standard of practice.

BE RESPONSIBLE AND ACCOUNTABLE

Principle A requires ethical statistical practitioners to take full responsibility for their work and support valid and prudent decision-making with appropriate methodology. As statistical practice is evolving toward delegating more work and even decision-making to AI, it is increasingly critical to emphasize that statistical practitioners (not AI) should assume responsibility, oversee automated solutions, and maintain their accountability when incorporating AI in statistical practice. They can reinforce this by implementing model governance, which involves setting up quality control review points, monitoring the models over time, and formulating a plan to cope with possible review outcomes. Other efforts include clearly defining operating constraints and being open about what statistical operations can and cannot be done with AI.

LEARN MORE

The *Nature* article titled "The Global Landscape of AI Ethics Guidelines" notes the growing global convergence of AI ethics around five ethical principles: transparency; justice and fairness; non-maleficence; responsibility; and privacy. Find it at *https://go.nature.com/3OPFFNG*.

A report by the Royal Statistical Society and Institute and Faculty of Actuaries, titled "A Guide for Ethical Data Science," addresses the ethical and professional challenges in AI-assisted data science practice. Read it at *https://bit.ly/440PNHM*.

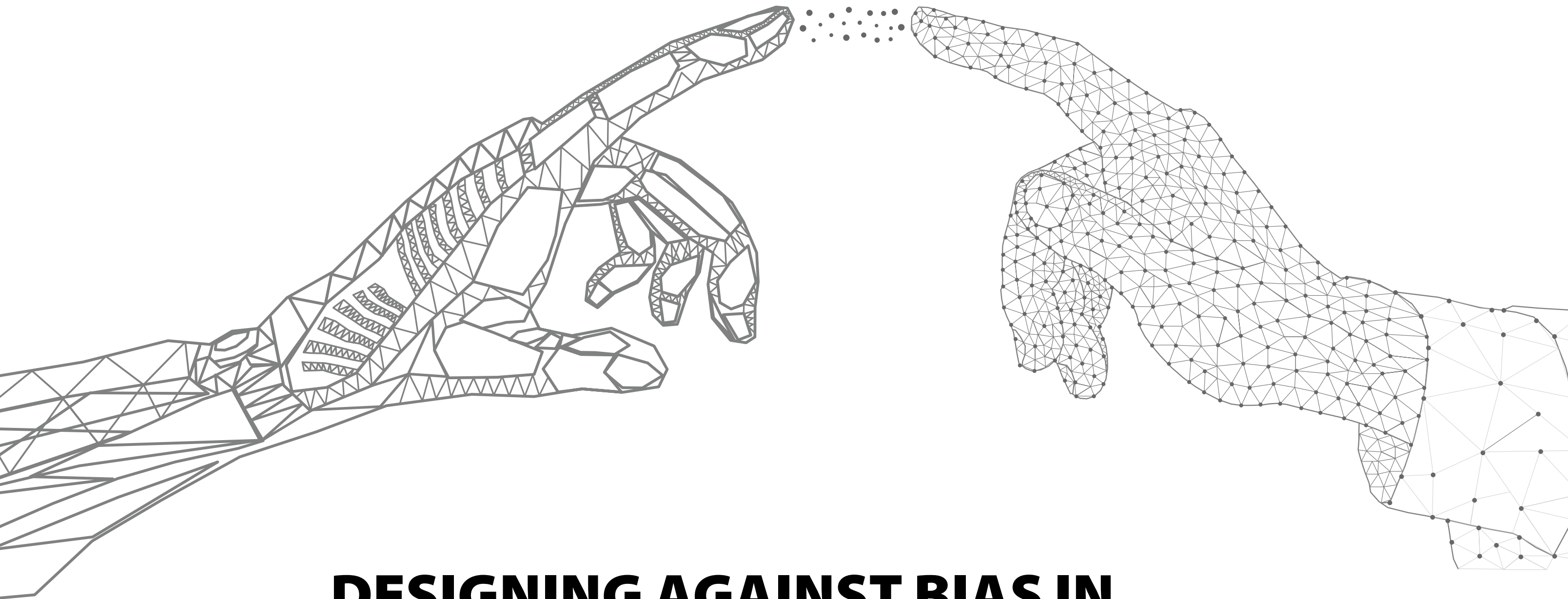
TAKE A PROACTIVE ROLE IN AI DEVELOPMENT IN STATISTICAL PRACTICE

Principle B in the ASA Ethical Guidelines states ethical statistical practitioners should develop and implement plans to validate assumptions and assess performance over time for models and algorithms designed to repeatedly inform or implement decisions. Promoting interpretable and transparent AI tools can make AI applications more trustworthy.

Statistical practitioners have a certain advantage in proposing transparent and rigorous modeling approaches. They can help build a more ethical environment for AI implementation by working with researchers from other fields to promote interpretability and validation in AI development and deployment. For example, many AI algorithms only provide a deterministic outcome without a valid measure of uncertainty. Statistical practitioners are trained to add confidence estimates to the findings. Their expertise can be valuable in making AI implementation more informative and statistically sound.

IN A NUTSHELL

We recognize ethical issues in statistical practice in the AI era are complex and multifaceted. They arise from the development, deployment, and use of AI technologies that can have significant impacts on our society and the world. There are many challenges and uncertainties in translating ethical norms into practice. Ethical issues in statistical practice in the AI era require continuous attention, reflection, and dialogue among all parties involved. ■



DESIGNING AGAINST BIAS IN MACHINE LEARNING and AI

David Corliss

Bias in machine learning algorithms is one of the most important ethical and operational issues in statistical practice today. Shifting the focus to testing for and mitigating bias at the design stage instead of after code is released can help prevent many of the problems seen in ML and artificial intelligence.

Confirmation bias is common, resulting from the human tendency to acquire information, analyze it, and develop explanations that confirm preexisting beliefs. It can lead to the selective inclusion of data sources that produce the results expected. While this may seem obvious, it can be insidious in actual practice. Algorithm developers can easily dismiss data sources as untrustworthy after seeing the results, making it important to have independent review of the data included in a training set.

A biased training set unrepresentative of the population of interest can result from convenience sampling, where an algorithm is developed using all the people available or who chose to answer a survey. This has been seen in voice recognition systems, where female voices are much more likely to be misunderstood.

Not using over-sampling for smaller population subsets can result in poor model performance for those groups, which are often made up of minorities. Alex Najibi at Harvard found that several facial recognition programs from leading companies had a lower accuracy for women and persons of color.

Prejudice bias results from training an algorithm with data labeling taken from previous human decisions, teaching the algorithm to replicate the very human bias many are created to avoid. While known as prejudice bias in literature, the source of the problem is unscreened previous human decisions in the training set. It can also occur in instances in which no prejudice is involved, such as in AI used to improve identification of part defects in a manufacturing plant.

Another type of bias can occur when there is a huge number of predictors, such as in text analytics. One example is using AI to screen résumés, then losing points if the text includes the name of a school with a female student body because people from that school were seldom hired in the past. If candidate predictors are included without screening individual predictors for bias, biased predictors can appear in the final model.

Another problem affecting algorithm bias is lack of transparency in the model. While not a source of bias in and of itself, lack of algorithm transparency can greatly complicate testing, identifying and confirming potential bias, and mitigating the effects. When the features included in an algorithm are withheld from the people using the algorithm and the people whose lives are significantly affected by its use, bias becomes more difficult to identify and mitigate.

Bias can be measured using disparate impact or benefit on marginalized minorities. An excellent example is the 2016 study of the COMPAS recidivism algorithm by Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin of *ProPublica*. The study found the algorithm misidentified Black persons as high risk more often than white people. It also found white people were more likely to be misidentified as low risk than Black. A key feature was the examination of the off-diagonal elements of the confusion matrix. The producers of COMPAS pointed to the accuracy of the model—the percentage of correct predictions. However, the bias was found in the records the algorithm got wrong: Being a person of color was an important factor in overestimating the risk of recidivism.

Many metrics can be used to quantify bias in scientific studies. In my own experience, odds ratios are easier to explain to a nonscientific audience, including most legislators, agency managers, and the public.

In recent years, several statistical packages have been developed to measure algorithm bias. One of the best is the Fairlearn package, developed by a Microsoft team led by Miroslav Dudik and released for free public use in 2020. Fairlearn calculates all the statistics needed to quantify algorithm bias, including confusion matrix and odds ratios for disparate impact. It features visualizations to help users understand the amount of bias in different model variants and balance it against predictive strength.

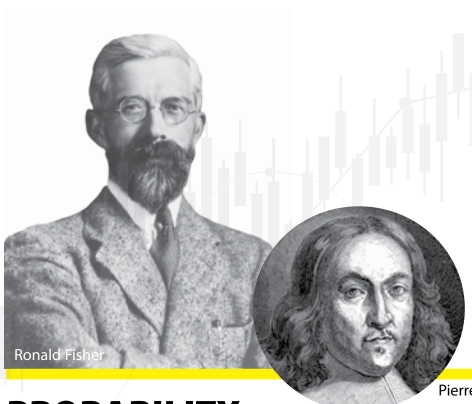
Understanding and avoiding the common sources of bias allows mitigation at the design stage—before an algorithm is released and subsequently found to have problems. New tools such as Fairlearn can be used to quantify and mitigate bias. The establishment and practice of standardized bias testing during design and development will result in AI that is more accurate, widely applicable, and fair to everyone. ■

David Corliss is the principal data scientist at Grafham Analytics. He also serves on the steering committee for the Conference on Statistical Practice and is the founder of Peace-Work.

ARTIFICIAL INTELLIGENCE, STATISTICS, and STATISTICIANS

Penny S. Reynolds, ASA History of Statistics Special Interest Group Member

Artificial intelligence is nearly always associated with computer science and engineering. Although obviously dependent on massive computational resources, AI has nevertheless required substantial statistical input along its entire developmental path. Here is an overview of major statistical concepts and methods integral to AI and the statisticians involved in their development.



Ronald Fisher

Pierre de Fermat

PROBABILITY

Probability is a fundamental concept in statistics. Modern AI is based on probability theory for quantifying uncertainty and making data-based forecasts. Development of the underlying mathematics during the 17th and 18th centuries was mostly motivated by the study of gambling. The cornerstone of modern probability theory was developed over the course of a year in letters between **Blaise Pascal** (1623–1662) and **Pierre de Fermat** (1601–1665).

Other foundational contributions included those by **Christiaan Huygens** (1629–1695) and **Abraham de Moivre** (1667–1754). **Pierre-Simon** and **Marquis de Laplace** (1749–1827) made perhaps the most important contributions to the mathematical theory of inference, particularly what today is recognized as a Bayesian interpretation of probability.

More than a century later, in 1922, **Ronald Fisher** (1890–1962) presented his unifying theory of statistics and inference. In this ground-breaking paper, he introduced several concepts central to statistics and AI such as consistency, efficiency, sufficiency, validity, likelihood, and—in particular—the statistical concept of information.

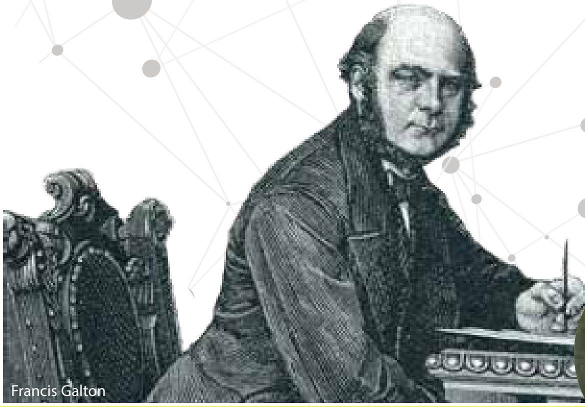
BAYESIAN

BAYESIAN STATISTICS

The Bayesian interpretation of probability is a description of the conditional probability of an event based on both observed data and prior information. Because they enable probabilities to be updated as new information becomes available, Bayesian statistics are instrumental in developing efficient machine learning algorithms and predictive models, as well as for decision-making under uncertainty. **Thomas Bayes** (1701–1761) is best known for his eponymous theorem, although he never published it. His notes describing a solution to one problem of inverse probability were published posthumously by **Richard Price** (1723–1791) in 1761.

Harold Jeffreys (1891–1989) and his 1939 book, *Theory of Probability,* was a major influence on the revival of Bayesian probability. During World War II, **Alan Turing** (1912–1954) developed Bayesian statistical methods as part of his Bletchley Park work in cryptanalysis. Although he does not explicitly use the term “Bayes,” he describes adjustment of the initial odds of a hypothesis by a prior and defines the expected Bayes factors against a true hypothesis.

Fisher redefined inverse probability as specifically “Bayesian” in 1950. In 1985, **Judea Pearl** (1936–) coined the term “Bayesian network” to describe models of conditional dependencies between sets of variables.


Francis Galton


Raymond Pearl


David R. Cox

REGRESSION

Regression methods are the backbone of machine learning and AI algorithms and remain the most common statistical applications for describing and predicting relationships between multiple variables. Feedforward neural networks and deep-learning algorithms are based on regression methods. **Francis Galton** (1822–1911) introduced linear regression in 1885 for quantifying relationships between variables. It was based on the method of least squares developed in the early 19th century by **Adrien-Marie Legendre** (1752–1833) and **Carl Friedrich Gauss** (1777–1855). In 1922, Fisher introduced the modern regression model. This effectively synthesized the concept of least-squares theory with the concepts of regression and correlation proposed by Galton, later expanded upon by **Karl Pearson** (1857–1936) and **George Udny Yule** (1871–1951).

Logistic regression has been described as the "go-to" method for machine learning involving classification of binary variables. The logistic function was initially developed between 1838 and 1847 in three papers by **Pierre François Verhulst** (1804–1849), a student of **Adolphe Quetelet** (1796–1874). Quetelet is probably best known for his concept of the "average man," the development of the body mass index, and his role as statistical mentor to Florence Nightingale.

The logistic function was repeatedly rediscovered, first by **Raymond Pearl** (1879–1940; ASA president, 1939) and **Lowell Reed** (ASA president, 1951) in 1920, then by Yule in 1925, who revived the name "logistic."

Joseph Berkson (1899–1982) was the primary developer of the modern logistic regression (he also coined the term "logit"). **David R. Cox** (1924–2022) further extended the logistic regression to models of observational data and developed its multinomial generalization.

Nonlinear regression is widely used in AI. It is an important tool for modeling data that is nonlinear in the parameters and thus poorly represented by linear models. Parameter estimates usually have no closed solution but are approximated by computationally intensive numerical optimization algorithms. The earliest of these was the Gauss-Newton algorithm, described by **Carl Friedrich Gauss** in 1809 as an extension of Isaac Newton's methods for determining the minimum of a nonlinear function.

Pafnuty Chebyshev (Tchebychev, Чебышёв 1821–1894) developed methods for polynomial series expansions for curve fitting; later, these were applied to descriptions of nonlinear dynamic systems.

The method of gradient descent was proposed in 1847 by **Augustin-Louis Cauchy** (1789–1857). Backpropagation with gradient descent is useful in neural network applications to improve prediction accuracy and error minimization.

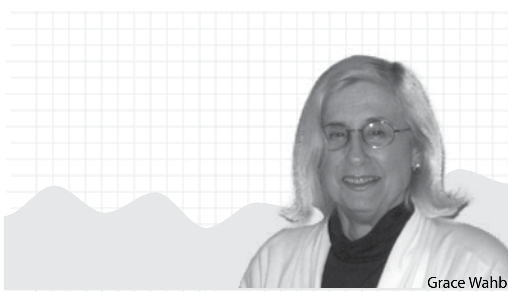
A more robust method than either Gauss-Newton or gradient descent is the Levenberg-Marquardt algorithm developed by statistician **Kenneth Levenberg** (1919–1973) in 1944 and rediscovered by **Donald Marquardt** (1929–1997; ASA president, 1986) in 1963.


Karl Pearson

Blaise Pascal

SEQUENTIAL ANALYSIS

Sequential analysis involves the process of data evaluation and decision-making in real time, with updating as more information is acquired. In essence, it involves the process of statistical estimation with sequential multiple hypothesis tests. It may have had its roots in the Gambler's Ruin problem formulated by **Christiaan Huygens, Blaise Pascal,** and de Fermat. The method has been attributed primarily to **Abraham Wald** (1902–1950), developed when he was working with the Columbia Statistical Research Group during World War II. Working independently, **Alan Turing** (1912–1954) developed similar sequential conditional probability analysis methods (Banburismus and Turingery) for decoding the German Enigma and Lorenz ciphers. This work remained classified until the early 1980s, so is not as well known.



Grace Wahba

SPLINE SMOOTHING

Spline fits are a relatively recent body of methods involving the fit of regression models to 'smooth' out noisy data and enable pattern recognition. **Isaac Jacob Schoenberg** (1903–1990) introduced the theory of splines in the 1940s.

The pioneering and enormously influential work of **Grace Wahba** (1934–) on smoothing spline functions, reproducing kernel Hilbert space theory, high-dimensional optimization, and generalized cross-validation has found wide application in the development of statistical machine learning, bioinformatics, medical imaging, computer graphics, and computer animation.

Bradley Efron

BOOTSTRAPPING

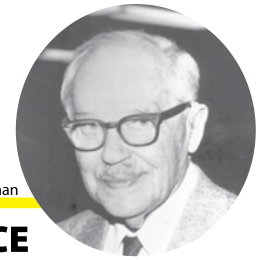
Bootstrapping was developed by **Bradley Efron** (1938– ; ASA president, 2004) in 1979 as a more versatile alternative to the nonparametric jackknife resampling method of **Maurice Quenouille** (1924–1973). Bootstrap resampling is a computer-intensive method for approximating the sampling distribution of almost any estimator. It has been called pioneering and hugely influential for its extraordinary versatility and applicability to a variety of disciplines, including AI.

Bootstrapping is also one of several techniques used for machine learning model validation. Models can be used to infer results for the population by 'training' on the bootstrapped data and then testing model predictions on external data sets.

PREDICTIVE MODELING, FEEDBACK, AND VALIDATION

Predictive analytic models based on machine learning or deep learning are used for classification, clustering, forecasting, and anomaly detection. An early nonparametric supervised machine learning method is the k-nearest neighbors classification algorithm (k-NN) developed by **Evelyn Fix** (1904–1965) and **Joseph Hodges** (1922–2000) in 1951. The k-means clustering algorithm is a supervised learning method developed by **J.B. MacQueen** (1929–2014) in 1967.

For a more detailed exposition of the role of statistics and statisticians in the development of artificial intelligence, see "Is There a Role for Statistics in Artificial Intelligence?" by Sarah Friedrich, Gerd Antes, Sigrid Behr, and coauthors in *Advances in Data Analysis and Classification*.



Jerzy Neyman

CAUSAL INFERENCE

The biggest challenge for machine learning and AI algorithms is distinguishing causation from correlation. **Jerzy Neyman** (1894–1981) has been credited for providing the earliest formal notation defining causal effects in 1923. Earlier, in 1921, **Sewall Wright** (1889–1988) developed path analysis to describe patterns of directed dependencies among a set of variables. Although these models could not explicitly determine causality *per se,* path analysis was an important precursor to structural equation modeling.

Pearl considers path analysis to be directly ancestral to causal inference methods. Pearl, himself, has been called "one of the giants in the field of artificial intelligence."

Ada King, Countess of Lovelace

John Tukey

COMPUTERS AND COMPUTATION

The idea that mechanical devices can generate new knowledge—generative AI—is surprisingly old. In the 1726 satirical novel *Gulliver's Travels,* Jonathan Swift describes a "wonderful machine" that would automatically generate books on all the arts and sciences "without the least assistance from genius or study."

Ada King, Countess of Lovelace (1815–1852), often credited as the earliest computer programmer, speculated on what she called "a calculus of the nervous system." This was the application of mathematical models to understanding thought and emotion—the first glimmers of the idea of a neural network. However, she apparently discounted the idea of artificial intelligence as such, concluding machines could not develop the capacity to "originate anything" or have the "power of anticipating any analytical relations or truths."

This was disputed in the epochal 1950 article "Computing Machinery and Intelligence" by Turing. Widely regarded as the father of artificial intelligence, Turing explicitly posed the question, "Can machines think?" He proposed the Turing test as the definition of the standard for an "intelligent" machine. He developed key ideas related to modern computing and programming (the hypothetical Universal Turing Machine) almost a decade before the technology was sufficiently advanced to put them into practice.

Claude Shannon (1916–2001), developer of information theory and founder of the digital revolution, independently developed many ideas like Turing's. He performed some of the earliest experiments in artificial intelligence, including the development of a maze-solving mechanical mouse and a chess-playing computer program.

Artificial intelligence applications rely heavily on interactive data visualization tools. **John Tukey** (1915–2000), the Father of Data Science, pioneered numerous statistical methods for computer application. In 1972, he devised PRIM-9, the first interactive dynamic computer graphics program for the exploration and identification of patterns in multivariate data. PRIM-9 has been described as revolutionary for its emphasis on the computer-human interface at a time when statistics was widely taken to be synonymous with inference and hypotheses testing. ■

JEDI CORNER

The Math Alliance

David Goldberg, Jacqueline Hughes-Oliver, Leslie McClure, and Javier Rojo

The Justice, Equity, Diversity, and Inclusion (JEDI) Outreach Group Corner is a regular component of Amstat News *in which statisticians write about and educate our community about JEDI-related matters. If you have an idea or article for the column, email the JEDI Corner manager at* jedicorner@datascijedi.org.

All authors of this month's JEDI Corner are leaders and mentors in the Math Alliance. David Goldberg is the executive director; Jacqueline Hughes-Oliver is associate director of statistics and a member of the executive council; Leslie McClure is on the executive council and served as its chair for three years; and Javier Rojo is associate director for biostatistics and data science.

The Math Alliance is a national mentoring community of faculty (mentors) and students (scholars) focusing on increasing traditionally excluded American minorities in the quantitative sciences professions. It began in the mid-1990s as an effort by the University of Iowa Department of Mathematics faculty to increase the number of underrepresented PhD students in the department. The project grew into a consortium among the three Iowa Regents Universities and four historically Black colleges and universities and was supported by an Infrastructure Grant by the National Science Foundation Division of Mathematical Sciences. In 2007, it transitioned to a national program and, in 2016, moved its administrative home to Purdue University.

Since its beginning, statistical sciences has had a strong presence in the Math Alliance. The late Kathryn Chaloner was a co-director and member of the board of directors. She began the Math Alliance Statistics Initiative to ensure statistical and data sciences were always well represented and to introduce scholars to statistics.

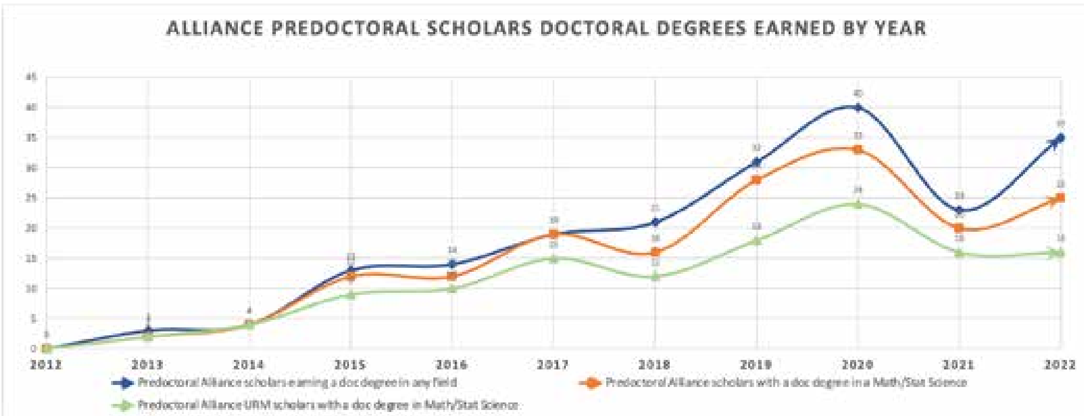
Several mechanisms within the Math Alliance facilitate students' transition into doctoral programs and the professions. Mentors are responsible for identifying and nominating Math Alliance scholars, thus every student comes to the alliance with a mentor. Graduate programs can be certified as graduate program groups, indicating a welcoming and supportive environment for Math Alliance scholars. There are now 42 doctoral program groups and 11 master's program groups, which also demonstrate clear paths to doctoral studies. At least 16 doctoral and four master's program groups grant statistical science degrees. There is also a regional structure, with autonomous organizations providing community and activities to a wide range of students and institutions in specific geographical areas.

Academic institutions, professional organizations, and corporations can become partners in the administrative Center for the National Math Sciences Alliance. Bristol Meyers Squibb recently became a partner, though the company has been engaging the community with a monthly "hot topics" session for the last couple of years. These sessions feature professionals with a wide range of expertise presenting insights into career options.

The Math Alliance structure relies on mentors to engage students and nominate scholars. Mentors sign up on the Math Alliance website. Currently, the alliance needs mentors from outside academia and is developing a mentoring corps from government, industry, professional organizations, and business.

Scholars who are within two years of completing a BS, bridge, or terminal MS program can become part of the facilitated graduate admissions procedure, where they are matched with a mentor in a doctoral program (faculty facilitator). The scholar, facilitator, and scholar's nominating mentor then work as a team to navigate the graduate application process. These scholars can also participate in the Career Paths Workshop, a two-day meeting to help illuminate the myriad careers and disciplines for which strong quantitative training is considered excellent preparation.

The entire Math Alliance community comes together for the annual Field of Dreams



MORE ONLINE
View a list of Center for the National Math Sciences Alliance partners at *https://bit.ly/3DQyyOB*.

Learn more about the Field of Dreams Conference at *https://bit.ly/3OtG6vB*.

Attendees pick up their materials at the Field of Dreams Conference Graduate Fair registration table.
Photo by Donald Cole

By the Numbers

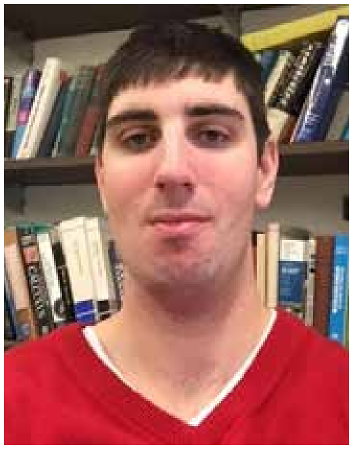
As the Math Alliance has grown, so has its success. When the Math Alliance became a national entity in 2007, there were 17 mentors and 59 students. Currently, the community has grown to 1,376 mentors at more than 410 colleges and universities. Of these mentors, 292 describe their research as involving statistical or data science, including 213 in doctoral programs.

There have been more than 2,500 Math Alliance scholars since 2007. Of scholars disclosing race or ethnicity, 80 percent self identify as an underrepresented minority (as defined by the National Science Foundation).

To date, at least 208 Math Alliance predoctoral scholars have earned doctorates. This includes at least 31 doctorates in statistical sciences. More than 900 students have participated in the facilitated graduate admissions procedure; 232 of those scholars are currently in doctoral programs. At least 56 scholars who participated in the facilitated graduate admissions procedure have earned doctorates, including at least eight in the statistical sciences.

Conference, which includes about 200 undergraduate and terminal MS students, along with current doctoral students; Math Alliance mentors; and representatives from graduate programs, professional societies, and corporations. The conference features a series of panels that provide critical information about selecting and thriving in a graduate program. There are community-building activities for mentors and scholars, three plenary sessions (including the Kathryn Chaloner Memorial Lecture in the Statistical Sciences and their Applications), a fair for students to connect with internships and research experiences for undergraduates, a graduate program fair, and a career fair for Math Alliance scholars who have earned doctorates.

The alliance needs more mentors to find and engage students. To become a mentor, visit *https://bit.ly/450Y7sf*. ■



My ASA Story: Will Egan Biostatistician

I am a biostatistician at Vertex Pharmaceuticals in Boston, Massachusetts. I joined the pharmaceutical industry after completing my PhD in statistics from Purdue University in 2020. When I entered the graduate program at Purdue, I was encouraged to join the American Statistical Association by my then-department head, Rebecca Doerge. I naively joined as a student member, unaware of how critical ASA membership would be to my professional development.

For my undergraduate studies, I attended a small liberal arts college as a mathematics major. Fortunately, I had some terrific mathematics and astrophysics faculty who encouraged me to pursue graduate studies in statistics. In my senior year, I formed the Statistics Journal Club and struggled to create momentum. For graduate school, I sought out a large statistics department with applied research in Bayesian methods, stellar teaching opportunities, and a vibrant statistical consulting center. Purdue University was a natural fit.

At Purdue, I was fortunate to be mentored by many outstanding statisticians like my adviser, Bruce A. Craig. One of my professors, Mark Ward (a current member of the ASA Committee on Membership Retention and Recruitment), approached me about a new ASA initiative: student chapters. Eager to use Purdue's STEM focus and unite both undergraduate and graduate students, I founded the Purdue ASA Student Chapter, serving as its inaugural president. It quickly became the largest in the country. To receive support from Purdue and the ASA was encouraging.

Through my student chapter experience, I became more involved in the ASA. First, I served on a working group on statistics education. As a graduate student, I was able to lead the Purdue team in the student leadership competition. We were coached by former ASA President Barry Nussbaum. Our team worked on data-driven membership questions and co-won the competition. The ASA then opened more opportunities for me. I achieved my GStat accreditation, participated in the Career Service at the Joint Statistical Meetings, and won an ASA Biopharmaceutical Section scholarship.

Building upon my experience in statistical organizations, I was honored to be invited to join the Committee on Membership Retention and Recruitment as a graduate student. I have served for several years and now chair the committee. I truly enjoy conveying the benefits of the ASA to the larger membership and helping to brainstorm about how to recruit new members.

My proudest ASA moment was successfully organizing a JSM 2022 invited session. Long had I tried in different forms to make the highly competitive invited session program, but I found a topic that really interested me. Encounters with my peers at Purdue convinced me data science was

There are more people to meet, sessions to organize, papers to read and write, and conferences to attend in the future chapters of my ASA story.

transforming the career interests of ASA members, particularly among students and early-career members. Additionally, academic departments began renaming themselves and methodological and computation research questions shifted to the fore. A recent membership survey conducted by a peer organization confirmed the growing interest in data science. Ultimately, I wanted to understand what exactly data science was (in contrast to statistics) and what this meant for the future of statistical organizations.

I reached out to several top statisticians and data scientists at various career levels and from different sectors. They enthusiastically joined. As JSM 2022 was my first major in-person event since the emergence of the COVID-19 pandemic, I was concerned attendance would be low. To my delight, the room was packed and the discussion vibrant. Even better, our session chair, Xiao-Li Meng, gave a prize for the best question. The winner was a high-school student, a rare feat for the invited session program at such a large conference.

Building on the success of JSM 2022, I co-organized and chaired a JSM 2023 invited session more closely related to my day-to-day work in the pharmaceutical industry, titled "Estimating Heterogeneous Treatment Effects to Inform Decisions."

My ASA Story is far from complete, and I look forward to future work on the Committee on Membership Retention and Recruitment. For my own professional development, I look forward to applying for my PStat accreditation when ready. There are more people to meet, sessions to organize, papers to read and write, and conferences to attend in future chapters of my ASA Story. ■

Biometrics Section News

The Biometrics Section and Mental Health Statistics Section recently chose Felicia Simpson as winner of the 2023 Annie T. Randall Innovator Award.

Simpson is an associate professor of statistics and chair of the department of mathematics at Winston-Salem State University. She is recognized for her outstanding contributions to statistical methods, innovation in education, and dedication to building a diverse profession through leadership and service. Her nomination highlighted her "cutting-edge and impactful" statistical research in gerontology and her role as an "innovator in expanding opportunities for underrepresented students to enter careers in statistics." Simpson received the award during the Biometrics Mixer during the Joint Statistical Meetings on August 7.

Winners of the 2023 Byar Award and student paper awards are the following:

Byar Award

- **Sarah Lotspeich,** Vanderbilt University, "Optimal Multi-Wave Validation of Secondary Use Data with Outcome and Exposure Misclassification"

Student Paper Awards

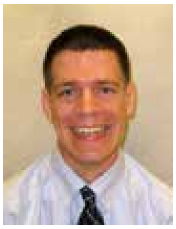
- **Rong Ma,** University of Pennsylvania, "A Spectral Method for Assessing and Combining Multiple Data Visualizations"
- **Cole Manschot,** North Carolina State University, "Interim Monitoring of Sequential Multiple Assignment Randomized Trials Using Partial Information"
- **Chao Cheng,** Yale University, "Doubly Robust Estimation and Sensitivity Analysis for Marginal Structural Quantile Models"
- **Bangyao Zhao,** University of Michigan, "Bayesian Learning of COVID-19 Vaccine Safety while Incorporating Adverse Events Ontology"

Awardees were also honored during the Biometrics Mixer on August 7 and gave presentations during the Biometrics Section Early Career Paper Award session on August 8.

STAT*tr@k*

On-Ramps for Data Science Experiences

Mark Daniel Ward



Mark Daniel Ward is a professor of statistics and (by courtesy) agricultural and biological engineering, computer science, mathematics, and public health at Purdue University. He is director of the Data Mine.

As the 2023–2024 academic year starts at colleges and universities all over the country, I reflect on ways mentors can create innovative, meaningful experiences for students. Such experiences not only include research guided by a faculty member but also research experiences in which students are working directly—across disciplinary boundaries—with mentors from industry.

In this regard, I am biased because our undergraduate and graduate students in the Data Mine worked on more than 80 projects during the 2022–2023 academic year, and it appears this year will be even more exciting for our team.

The Data Mine is a program in the Office of the Provost at Purdue University that provides students an on-ramp into the data sciences, even if they do not have a background in data science. We are projecting an enrollment of 1,700 undergraduate and graduate students this fall, not only from Purdue but also from across the United States.

In addition to our program on campus, we coordinate the Lilly Endowment–funded Indiana Data Mine, the ASA's National Science Foundation–funded National Data Mine Network, and the NSF-funded Developing Experiential Accessible Framework for Partnerships and Opportunities in Data Science for the deaf community (DEAF PODS).

I think students in data science and statistics appreciate opportunities to collaborate with students from other areas. Similarly, students from other disciplines (e.g., business, engineering, health science, technology, etc.) value the opportunity to gain insight from statistics and data science students, who can develop statistical or machine learning models.

Unlike a classroom experience, these interdisciplinary experiences are a more accurate mirror of the types of environments in which students will work when they have full-time careers.

Brad Fruth, director of innovation at Beck's Hybrids, talks about working with students in a video at *www.youtube.com/watch?v=dXferJvntko*.

Photo courtesy of Purdue Marketing and Communication

Students seem to enjoy learning about a domain from an industry practitioner. There is simply no substitute to learning directly from a person who has worked in a domain throughout their life. Such mentoring is especially meaningful if that person is willing to share their experiences with students throughout a nine-month partnership.

If a student can meet with colleagues from industry one or two times a week throughout an academic year, I wager such discussions will be one of the student's most worthwhile experiences in college. Even the most experienced statistics and data science faculty are usually unable to replicate the types of advice a colleague from industry can share with students over an extended period.

As faculty members, I think we should work hard to incorporate deep, rich interactions with industry experts throughout the undergraduate and graduate experience.

I think students are more likely to understand AI-powered tools if they are involved in developing them (rather than, say, only studying the mathematical frameworks or computational aspects). By being included in an AI research project, students will not only feel a sense of ownership but a sense of belonging and self-efficacy. Moreover, students who build AI models and have an early first encounter with AI in their career will also have more time to consider the many ways AI is transforming companies for which they may work after college or graduate school.

Another aspect of this alignment of students' careers with statistics and data science is broadening career pathways. By working with a company early in one's college experience, a student has the opportunity to develop a first-hand appreciation for the mission of a small or midsize business, or with a government agency, where they would (otherwise) never have considered working. In states like Indiana, where the brain drain to coastal states can be a concern, this is a crucial economic issue.

My team members at the Data Mine are delighted when a student chooses, for instance, to work with Beck's Hybrids (a family-owned seed company) or the Indiana Family and Social Services Administration. (The full list of partners is available at *https://datamine.purdue.edu/symposium*).

Brad Fruth, director of innovation at Beck's Hybrids, talks about working with students in a video at *www.youtube.com/watch?v=dXferJvntko*. One project he discusses focused on optimizing the supply chain at Beck's Hybrids and another focused on statistical and data-driven models for choosing test plot locations.

At the end of the video, the interviewer asks, "Would you recommend other corporations or companies to partner with the Data Mine and why?" Fruth responds tongue-in-cheek, "No! Because we want the students for ourselves, and selfishly, this is all ours for world domination. Don't do it!"

Michael Douglass, program engineer at Raytheon, also loves working with undergraduate and graduate students on real-world research problems. He has repeatedly told our team working at the Data Mine is now a condition of his employment. In other words, he told his supervisor he will quit if he is unable to continue working with the students. Douglass' mentoring goes far beyond team meetings with the students. He goes on long bike rides with them and often has meals with them in the dining court of their residence halls. He truly understands the deep impact of mentoring students early in their careers.

Students in the Data Mine are thankful for these experiences. As I was writing this article, first-generation university student Taylor Saunders stopped by my office to discuss her plans for graduate school. When setting up our meeting, she was gushing about her research experience with our team this summer, saying, "I have learned more than I expected and have grown both professionally and personally." She emphasized she has "found the experience to be both vigorous and exhilarating."

Saunders is working with a team of peers in our NSF-funded DEAF PODS program. She is studying at Arizona State University but spending the summer at Purdue. Her research project is provided by the Indiana Family and Social Services Administration.

In addition to the mentoring provided by our team, Cristian Guandique, deputy director of data science and engineering, and Matt Kirby, director of engagement and analytics, at the Indiana Family and Social Services Administration meet regularly with the students—both online and at Purdue—to guide their research. This experience enables Saunders and her teammates to gain an immediate understanding of how statistics and data science are practiced in the state government to provide a multitude of services for people in Indiana.

After Saunders completes her summer work with DEAF PODS, she plans to join the ASA's National Data Mine Network. She wrote to say the following:

> I am extremely appreciative to have been given this opportunity, and as of today Jessica [Jud, the Data Mine Senior Manager of Expansion Operations] connected me to [the ASA's] National Data Mine Network. I am extremely thankful for all of the hard work you and your team put into the Data Mine every day to make these opportunities for myself and other students possible. I spoke to David Glass [the Data Mine managing director of data science] today about further opportunities and mentioned my desire to earn my master's degree after my graduation date. The Data Mine, and the community here at Purdue, has really inspired me to further invest in my education and skills after my bachelor's, and I expressed interest in Purdue's statistics program[…] I am extremely passionate about data science and this program has opened my eyes to all of the possibilities ahead of me.

For faculty developing interdisciplinary applied data science programs, the amount of background in computation, mathematics, and statistics needed is many times a key question. In the Data Mine, we have taken the path of allowing students to join early in their studies, without having a background in these topics. We enable the students to learn data science competencies as they work on their research.

However, the amount of background needed for data science is a topic of ongoing debate. On July 13, for instance, Rob Gould weighed in on this topic in a *New York Times* article titled "In California, a Math Problem: Does Data Science = Algebra II." Additionally, a team of eight University of California faculty wrote a letter stating data science can "harm students from such groups by steering them away from being prepared for STEM majors." These discussions about what training is most appropriate or necessary for data science programs will likely continue for the foreseeable future.

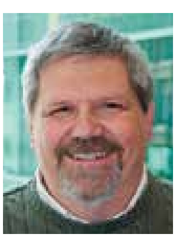
I want to emphasize there are many excellent student opportunities in statistics and data science research throughout the US. For instance, Talitha Washington is doing innovative work as the director of the Atlanta University Center Data Science Initiative. She is also co-PI of the ASA's National Data Mine Network, which has doubled its number of applicants from last year. This demonstrates the broad appetite for data science research experiences among students is increasing nationwide, including among students pursuing their college degree at minority-serving institutions.

Also, Sat Gupta and his colleagues at The University of North Carolina at Greensboro offer research experiences for undergraduates. Their research program was launched through an ASA initiative funded by the NSF and continues under the auspices of the university's own NSF research experiences for undergraduates grant.

I firmly believe early research experiences are some of the most valuable ways faculty and industry mentors can support students. Such experiences also give companies and universities new ways to build innovative partnerships. If you are not yet working with students on data science or statistics research experiences, I encourage you to think about exploring such opportunities. ■

STATS4GOOD

Ethical AI Groups Spring Up Around the World



David Corliss is the principal data scientist at Grafham Analytics. He serves on the steering committee for the Conference on Statistical Practice and is the founder of Peace-Work.

Ask any bot—ethical AI is a big deal.

The development and use of AI arguably goes back to the 1940s and code-breaking devices. Recently, as powerful new tools such as large language models have gained public visibility, there has been tremendous growth in the concern about AI and its capabilities, uses, and abuses. As a result, ethical AI teams are becoming firmly established in academia, industry, and government. With more AI applications being released every day, #DataForGood practitioners in every area are leading the way in developing and promoting ethical best practices.

A great review article from the producers of the Open Data Science Conference, titled "The AI Ethics Boom: 150 Ethical AI Startups and Industry Trends," provides a plethora of information about and analysis of industry trends. The article highlights the Ethical AI Database, listing more than 140 ethical AI startups in both commercial and NFT spaces.

A number of independent not-for-profit organizations were first responders to growing concerns about the ethics and effects of artificial intelligence. Distinguished ethical AI leader Timnit Gebru has been a driving force in this space, founding the Distributed Artificial Intelligence Research Institute for community-centered AI research independent of the industry. He also founded Black in AI for Black professionals in artificial intelligence.

The number of organizations continues to grow rapidly. The post-pandemic technology landscape makes it easy to get involved with a group working in your particular area of interest, no matter where you and the ethical AI organization are located.

Major research universities have been fertile ground for ethical AI research. An early leader in this space is Stanford's Institute for Human-Centered AI (HAI). Established in 2019,

HAI supports research, educational classes and materials, and policy development. A focus at Stanford is the effect of AI on the human experience and society as a whole. HAI is active in hosting events, including webinars and a boot camp for congressional advocacy.

Oxford University's Institute for Ethics in AI works globally to connect ethicists, philosophers, and social scientists with academia, industry, and government. Influenced by the impact of philosophers on the establishment of ethical practices in medicine decades ago, the Oxford Institute fosters the establishment of ethics in AI as a standardized area of practice. They seek to have ethical training included as an integral part of tech training and are developing a program for visiting professors to participate in.

Developing partnerships with academia and industry is a focus of the University of Michigan Institute for Data Science. As a leading public research university, its mission includes a strong commitment to extending beyond academic research to serve intersectionality across a wide range of interests to the public. The institute recently hosted a one-day forum titled "From Theory to Practice: Building Ethical and Trustworthy AI," that could serve as a model for others. It featured presentations from Rocket Company's dedicated ethical AI team.

It's been my observation that industry tends toward creating a team within their IT organization while most university and government programs seek to be interdisciplinary, bringing together researchers from many departments. In response to this need, the University of Michigan Institute for Data Science is developing training materials that will help companies set up their own ethical AI teams.

I asked ChatGPT if ethical AI is important. Its response was a page and a half long, and I didn't read it before finishing this column to avoid any influence on my writing. Given how LLMs work, the long, detailed response shows a lot of work is being done in this area. ChatGPT has a list of limitations on its home page, including "May occasionally generate incorrect information. May occasionally produce harmful instructions or biased content. Limited knowledge of world and events after 2021." This shows the impact ethical AI is having on the tools and how they are presented to the public, reflecting concerns about how the results might be used.

As the use of AI grows—along with increasing public awareness of the ways it affects our lives every day—ethical AI will become even more important. There are myriad ways Data for Good practitioners can get involved, including while doing independent research or working for a university, not-for-profit, or ethical AI team at a tech-driven company. Whatever your area of work, ethical AI is now part of it as an essential element of Data for Good. ■

Getting Involved

In Data for Good opportunities, this is the month for *Amstat News'* special issue on the role statistics and statisticians play in AI. Check out the whole issue to find new ideas, inspiration, and resources for your next D4G project.

We also want to highlight the Black in AI workshop in December (*https://bit.ly/45iqe6x*). This event will be hybrid and co-located with the NeurIPS conference on neural information processing systems. Check out the conference website at *https://nips.cc* to get all the details and share it with a friend.

PROFESSIONAL OPPORTUNITY LISTINGS
To advertise, visit
ww2.amstat.org/ads.

New York

■ Applications are invited for a tenure-track position in statistics at Vassar College to begin fall 2024; see the full ad at *https://apptrkr.com/4402982*. The ideal candidate is committed to excellence in scholarship and undergraduate teaching and working with the department's three statisticians to further expand statistics and data science at Vassar. Application review will begin October 2, 2023, and continue until the position is filled. Vassar College is deeply committed to increasing the diversity of the campus community and the curriculum, and to promoting an environment of equality, inclusion, and respect for difference. Candidates who can contribute to this goal through their teaching, research, advising, and other activities are encouraged to identify their strengths and experiences in this area. The College is an Equal Opportunity and Affirmative Action employer, and especially welcomes applications from veterans, women, individuals with disabilities, and members of racial, ethnic, and other groups whose underrepresentation in the American professoriate has been severe and longstanding. ■

This month's Top 10 is the 'Top Ten **Rejected Statistics and Data Science T-Shirt Slogans**.'

Amstat News continues its hilarious offering from ASA Executive Director Ron Wasserstein. Each month, he delivers a special "Top 10"—one that aired during a recent edition of the *Practical Significance* podcast. Last month, Ron offered up the "Top Ten Statistics and Data Science T-Shirt Slogans" for folks to 'flex' at JSM 2023. This month, he follows up with the "Top Ten Rejected T-Shirt Slogans." He admits, "I'd wear most of these T-shirts!"

Wasserstein

PODCAST

To listen to the *Practical Significance* podcast, visit *https://magazine.amstat.org/podcast-2*.

10

Warning: I am the product of unsupervised learning.

09

Permanently in the rejection region

08

You come to me with this data now? You're kidding, right?

07

Approaching normal asymptotically

06

Data munging? Looks like data fudging to me.

05

No. Two percent of your grant time is not enough.

04

Visualize THIS! (You'll have to picture the raised finger for yourself.)

03

Proud to be uncorrelated with everything

02

I'm not only fit—I'm overfit.

#01

Your AI whispers my name in its sleep.