PRINCETON
School of Engineering
and Applied Science

Supplement to the *Princeton Alumni Weekly*

EQuad News

Spring 2024
Volume 35, Number 2

Photo by David Kelly Crow

Accelerating innovation with AI

With its uncanny ability to converse like humans, ChatGPT has ignited people's imaginations and concerns. At Princeton Engineering, we see AI as a powerful tool that can enhance human capability to achieve what humans could never do on their own. That is, after all, what inventors and engineers have done for millennia — build new technologies that enhance human capabilities and improve lives, from the wheel to spacecraft.

In this magazine, in addition to recent highlights from across Princeton Engineering, we show some of the many ways Princeton engineers are building the fundamental tools of AI and applying them to pressing societal challenges, such as the quests for clean energy, better medicines, new materials, and safer robots.

The key to these strengths — and the strategy behind major ongoing investments by Princeton — is our interdisciplinary approach. Each of our six engineering departments is using AI to accelerate their work, often in collaboration with each other and with the natural sciences, social sciences, humanities, and the arts.

This interdisciplinary spirit drives our growth in many areas, including bioengineering, quantum engineering, robotics, sustainability, and next-generation wireless networks. At Princeton our engineers work collaboratively with a relentless drive toward harnessing technology to benefit humanity.

Andrea Goldsmith
Dean | Arthur LeGrand Doty
Professor of Electrical
and Computer Engineering

EQuad News
Spring 2024
Volume 35, Number 2

Dean
Andrea Goldsmith

Vice Dean
Antoine Kahn *78

Associate Dean, Undergraduate Affairs
Peter Bogucki

Associate Dean, Development
Jane Maggard

Associate Dean, Diversity and Inclusion
Julie Yun

Director of Engineering Communications
Steven Schultz

Senior Editor
John Sullivan

Digital Media Editor
Aaron Nathans

Content Strategist
Molly Sharlach

Senior Engineering Communications Strategist
Scott Lyon

Communications Specialists
Julia Schwarz
Wright Señeres

Development and Communications Coordinator
Christine Fairsmith

Contributors
Liz Fuller-Wright
Adam Hadhazy
Colton Poore

Graphic Designer
Matilda Luk

Web Designer
Neil Adelantar

EQuad News is published twice a year by the Office of Engineering Communications in collaboration with the Princeton University Office of Communications.

EQuad News
C-222, EQuad,
Princeton University
Princeton, NJ 08544
T 609 258 4597
F 609 258 6744
eqn@princeton.edu

https://engineering.princeton.edu/news/equad-magazine

Recent Highlights

Planning for Princeton's NextG corporate affiliates program began at its inaugural symposium in 2023. Kaushik Sengupta, right, and Yasaman Ghasempour, center, are both co-directors of the initiative. *Photo by Sameer A. Khan/ Fotobuddy*

PRINCETON RESEARCHERS, INDUSTRY LEADERS DRIVE NEW ERA OF INNOVATION IN WIRELESS AND NETWORKING TECHNOLOGIES

Princeton's School of Engineering and Applied Science has launched a corporate partnerships program to bolster innovation and research in the next generation of intelligent wireless and networking technologies.

Part of Princeton's NextG Initiative, the membership-based program aims to increase the flow of knowledge and innovation between industry and academia and to engage policymakers to enable the greatest impact of these advances. It has attracted close to a dozen corporate members, including several of the world's leading networking and semiconductor companies. The companies will work with Princeton students and faculty members to ensure that the knowledge created by academic research finds its way out of labs and into practice, maximizing its benefit to society.

"The best academic research to help realize the promise of future wireless networks is informed by close interactions between faculty and industry technology leaders," said Andrea Goldsmith, dean of Princeton Engineering. "Our NextG program aims to foster those deep collaborations around wireless technology, as well as policy, that are required to drive meaningful innovation and global leadership in an era of rapid change."

The program's inaugural members include American Tower, Crown Castle, Ericsson, Intel, InterDigital, MediaTek, Nokia Bell Labs, Qualcomm Technologies, Samsung Research America, and Vodafone.

The initiative draws on Princeton's expertise across a range of research areas, including cloud and edge networks, intelligent sensing, terahertz and sub-terahertz communication, security and privacy, network resilience, and a host of applications from health care to transportation to defense.

Yasaman Ghasempour, an assistant professor of electrical and computer engineering, said that while this initiative has grown organically from the wealth of interdisciplinary research on campus, the corporate partners' insights, expertise, and real-world experience are instrumental in shaping the program's future activities.

"In academia, we tend to look at fundamental research questions that require long-term investment," Ghasempour said. "Industry develops technologies that can be realized and have applications in the real world today, or at least very soon. But there are a lot of synergies in the topics and challenges we tackle."

To learn more about becoming a member of Princeton's NextG Corporate Affiliates Program, contact Jessie Skinnider from the Office of Corporate Engagement and Foundation Relations at jskinnider@princeton.edu.
– by the Office of Engineering Communications

Students adapted a legendary speedboat, named Big Bird, for electric propulsion. *Photos by Princeton Electric Speedboating*

PRINCETON'S STUDENT TEAM BROKE THE WORLD SPEED RECORD FOR ELECTRIC BOATS

Averaging 114 mph, a team of Princeton students smashed the world speed record for electric boats on Oct. 26, 2023, surpassing the old record by 26 mph.

"Phenomenal," said team captain Andrew Robbins, a junior majoring in mechanical and aerospace engineering.

The Princeton team set the new record on a 1-kilometer course on Lake Townsend outside Greensboro, North Carolina. Piloted by veteran speedboat captain John Peeters, the Princeton boat made one pass through the course at 111 mph. Without recharging the batteries, Peeters made a return pass at over 117 mph. The average speed of 114 mph surpassed the previous record of 88 mph set by the automotive company Jaguar in 2018.

Students performed much of the work on the boat in space at the Forrestal campus.

Princeton Electric Speedboating, a team of 44 undergraduate and graduate students, represents most engineering disciplines as well as members studying economics and physics. The students conducted much of the engineering, design, and fabrication of the boat and of the engine that propelled it to the record-setting run.

Luigi Martinelli *87, the team's faculty adviser and a professor of mechanical and aerospace engineering, said Princeton Electric Speedboating is completely run by students. An expert in computational fluid dynamics and aerodynamic design, Martinelli said that while faculty members offer advice, the students manage the club's organization, engineering, and finances.

The Princeton team broke the record using a 14-foot outboard boat called Big Bird. With a hydroplane hull resembling a stealth fighter plane skimming over the water, Big Bird is equipped with a three-phase, 200-horsepower electric motor designed by the Princeton team and built in coordination with Flux Marine of Rhode Island.

The team also took first place in the 2023 national collegiate competition Promoting Electric Propulsion, organized by the American Society of Naval Engineers and sponsored by the U.S. Office of Naval Research.

Robbins, who piloted the boat in that competition, said team members were thrilled with both results. He said they plan to make further adjustments to the boat and improve on the results in the future.
– **by John Sullivan**

CLIMATE CHANGE LIKELY TO DRIVE MORE FLOODS IN SOME PARTS OF THE U.S., FEWER IN OTHERS

By condensing flooding analysis into its primary mechanisms, researchers at Princeton have projected that climate change will markedly impact river basin flooding across the United States during the 21st century.

In an article published in Nature Communications, the researchers found that projected variations in temperature and precipitation are expected to drive increased flooding in the Northeast and Southeast, particularly along the Eastern Seaboard. Flooding will generally decrease in the Southwest and the Northern Great Plains in areas including Montana and the Dakotas.

Corresponding author Gabriele Villarini said that by basing projections on mechanisms such as temperature and precipitation, the researchers took a new approach to flood analysis. He said most analyses examine the historical record and look for trends in data moving through the present and into the near future. Planners often refer to floods using the historical record, talking about a 100-year flood or a 1,000-year flood as a measure of severity. But Villarini said this type of analysis is insufficient during a changing climate.

"When we design protective structures, we design for the future, not for the past," said Villarini, a professor of civil and environmental engineering and Princeton's High Meadows Environmental Institute. "If all you do is look at the past and assume the future is just the same as what happened before, you will run into potential issues because of climate change."

Rather than rely only on the record, the researchers focused on understanding the processes that led to the year-to-year changes in flooding, and broke causes of flooding down into the two most important climate factors: seasonal temperature and precipitation. The researchers did extensive modeling on how each seasonal component affected flooding.

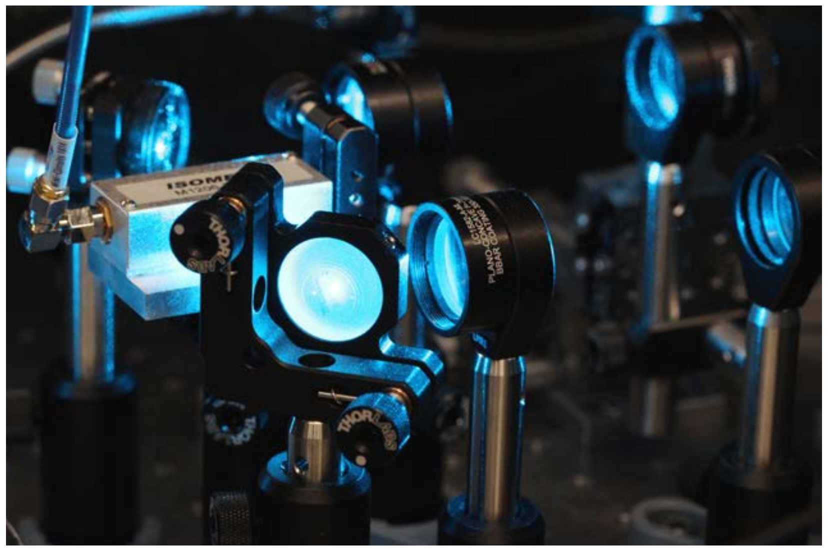
Flooding in Fenton, Missouri, in 2008. *Photo by Jocelyn Augustino, FEMA*

After validating the flooding model and the suitability of global climate models in reproducing the past, the researchers then turned to future events. They chose different scenarios for climate change-driven precipitation and temperature across the United States and used those to estimate changes in regional flooding. – **by the Office of Engineering Communications**

ILLUMINATING ERRORS CREATES A NEW PARADIGM FOR QUANTUM COMPUTING

Researchers have developed a way to reveal errors in quantum computers in real time, a significant step toward building machines that can tackle the world's hardest computational problems. This is a new direction for research into quantum computing hardware, which typically seeks to lower the probability of an error occurring in the first place.

Physicists have been inventing new qubits — the basic components of a quantum computer — for three decades. Steadily, those qubits have become less fragile and less error-prone. But some errors are inevitable, and the central obstacle to the development of quantum computers is correcting these errors.



Researchers led by Jeff Thompson, associate professor of electrical and computer engineering, developed a technique to make it 10 times easier to correct errors in a quantum computer. At the heart of the computer, lasers control the states of individual atoms, manipulating information based on the extremely subtle physics of those atoms' energy levels. *Photo by Frank Wojciechowski*

To correct an error, however, you must find it. And in the strange world of quantum technology, the process of checking for errors typically introduces more errors, which must be found again, and so on.

Quantum computers' ability to manage those inevitable errors has remained mostly stagnant over that long period, according to Jeff Thompson, associate professor of electrical and computer engineering, who led the study. But where others saw limitation, Thompson saw opportunity.

"Not all errors are created equal," he said. Some are easier to manage than others. Converting difficult errors into manageable errors would dramatically improve the ability to correct them.

By using a new approach to storing the information, based on a deep understanding of atomic energy levels, the researchers were able to characterize the probability of errors occurring in an array of 10 atomic qubits.

Not only did they find extremely low error rates but, more importantly, using the new approach, the process of characterizing the error rate did not destroy the information stored in the atoms, as many other techniques do. Rather, taking a measurement caused the qubits with errors to emit a flash of light, while the qubits without errors remained dark and unaffected.

This process, unfolding in real time, converts the errors into an erasure error, long known to be simpler to correct than errors in unknown locations, Thompson said.

"What's nice about erasure conversion is that it can be used in many different qubits and computer architectures," Thompson said, "so it can be deployed flexibly in combination with other developments." – **by the Office of Engineering Communications**

Researchers Néhémie Guillomaitre and Xiaohui Xu with samples of a new type of recyclable hydrogel that can be cut and molded into different forms. *Photo by Dan Komoda*

REUSABLE AND RECYCLABLE, THIS NEW HYDROGEL SQUISHES THE OLD VERSION'S ENVIRONMENTAL IMPACT

Princeton researchers have created a new type of hydrogel that is recyclable, yet still tough and stable enough for practical use (and reuse).

As flexible networks of polymer chains suffused by water, hydrogels possess excellent properties including softness, elasticity, and biocompatibility. Accordingly, the squishy materials have already found widespread use as contact lenses and wound dressings. Hydrogels also hold great promise for drug delivery systems, agriculture, and food packaging, among other applications.

Unfortunately, conventional hydrogels pose environmental pollution problems because they cannot be effectively recycled or reprocessed. Hydrogels also degrade from long-term use. The researchers said these limitations derive from the materials' structure. Conventional hydrogels rely on chemical bonds for their firmness and ability to soak up water and other solvents. On a chemical level, these bonds are cross-linked, meaning that bonds form among different polymer molecules within the hydrogel. This cross-linking, characteristic of resins undergoing curing or rubber vulcanization, gives hydrogels flexibility and strength. But it also makes them extremely difficult to separate into components for recycling.

The study's lead author, Xiaohui Xu, and her colleagues took a new approach to building hydrogels. Rather than relying on chemical bonds to connect different polymers, the researchers decided to harness phase separation, a familiar phenomenon in which mixed liquids, such as oil and water, separate into components.

"Hydrogels offer tremendous societal benefit, but their lack of sustainability has loomed as a significant issue," said Xu, a postdoctoral researcher in the lab of Rodney Priestley, the Pomeroy and Betty Perry Smith Professor of Chemical and Biological Engineering at Princeton. "In this study, we have shown how taking advantage of phase separation can lead to new kinds of hydrogels that are durable and recyclable and still have good mechanical properties."
– **by Adam Hadhazy**

DATA DRIVES QUICKER, SAFER DECISIONS FOR RACE CARS AND ROBOTS

Race car drivers handle a unique set of challenges, aiming to win while steering through a shifting pack of opponents. In a project to train algorithms to master this high-stakes task, Princeton researchers are applying knowledge of how individual decisions influence group behavior in uncertain and changing environments — a model problem with myriad applications.

Naomi Ehrich Leonard ’85, the Edwin S. Wilsey Professor of Mechanical and Aerospace Engineering and chair of the department, has long used mathematical models to understand the remarkably fast, robust, and adaptable group decision-making in natural systems like swarms of honeybees and flocks of birds. She has applied these models in settings including a group of robots exploring the ocean depths.

Now, Leonard has teamed up with Jaime Fernández Fisac, an expert in safe robotics and human-centered autonomy, to improve safety and performance in racing. Their methods combine the mathematics of group dynamics with techniques from machine learning and control systems, and use data from real and simulated races. The goal is to improve coordination on multiple levels — between each car’s AI and its driver, and among all the cars on the track.

With support from the Toyota Research Institute, the team is building neural networks — multilayered algorithms inspired by the organization of neurons in the brain — that will use training data to tune opinion dynamics models that ensure split-second breaking of costly decision-making deadlocks.

“We want to make sure that the AI can convey its situational awareness quickly and clearly,” said Fisac, an assistant professor of electrical and computer engineering. “Transparency is critical when you deal with increasingly sophisticated driver assistance systems,” like current systems that warn of lane departures or vehicles in the driver’s blind spot, or even nudge the steering or tap the brakes to avoid a potential accident.

Fisac and Leonard seek to develop onboard AI systems that can rapidly infer and adopt drivers’ shifting priorities during a race, as well as warn them about hazards or directly intervene in the vehicle’s control, all without unduly distracting or, even worse, startling the human behind the wheel.

Racing is “uniquely dynamic and safety-critical,” Fisac said. “You’re really pushing the car to the limits of its operating envelope, and a single misstep can have catastrophic consequences.” – **by Molly Sharlach**

Maximizing benefits.

Avoiding harms.

Making artificial intelligence work for humanity.

Princeton engineers have been building the foundational tools behind machine learning and artificial intelligence for decades, dating back to the visionary thinking of computer science forefather Alan Turing *38. Today, that legacy has broadened into three key strengths. First, researchers in engineering and computer science continue to advance the underlying algorithms and mathematics of AI, while developing a more rigorous understanding of how these complex systems work. Second, researchers across disciplines are collaborating widely to apply these emerging techniques to practical problems, accelerating innovation in fields from sustainability to robotics to medicine to the humanities. Third, experts in these areas are working with colleagues across the social sciences and humanities to think deeply about and to improve the impacts of AI on society.

Stories on the following pages show just some of Princeton’s projects across these three areas and their many overlaps. Please follow us at **engineering.princeton.edu** or on social media for all the latest.

DIVERSIFYING AI THROUGH DATA COLLECTION AND CAREER DEVELOPMENT

by Molly Sharlach

Today's computer vision models are incredibly adept, but because models are often trained on images scraped from the internet, they can reflect biases and stereotypes about places or people.

A decade ago, as a Ph.D. student at Stanford, Olga Russakovsky was part of a team that developed ImageNet, a groundbreaking visual database. Now an associate professor of computer science at Princeton, Russakovsky works to expand computer vision capabilities, while also making these AI models more fair, accurate, and transparent.

Olga Russakovsky. *Photo by David Kelly Crow*

Her team's recent work includes an effort to achieve more complete representation of everyday objects from around the globe. Because these images will be critical for applications like robotics, agriculture, and health care, Russakovsky's team commissioned photos from around the world. By adding images of objects like houses, toothbrushes, and wheelbarrows, the project seeks to counterbalance image sets that heavily rely on images from North America and Western Europe. Vikram Ramaswamy, who earned his Ph.D. from Princeton last year and is now a lecturer in the computer science department, is the lead author of the project, called Geode.

"We're part of this giant ecosystem that's rapidly evolving," Russakovsky said. "I think we have to really engage with these social questions as we're building our models and advancing the state of AI. We have to step up."

For Russakovsky, increasing the diversity of AI creators by lowering barriers goes hand in hand with thinking carefully about the technology's design and applications. Along with advancing ImageNet, Russakovsky and Stanford professor Fei-Fei Li '99 (with educator and mathematician Rick Sommer) cofounded AI4ALL, a nonprofit that started in 2015 as a high school summer program at Stanford. AI4ALL has introduced thousands of students to AI technology and associated ethical and policy issues, and offers support for entrepreneurship and career development.

Russakovsky codirects an AI4ALL summer program at Princeton, where high school students from low-income backgrounds learn programming and AI basics and collaborate on research projects. She played a key role in the early 2024 launch of AI4ALL Ignite, a virtual accelerator program that helps undergraduates delve into AI in ways that go beyond their school curricula. With guidance from industry experts, students develop independent projects, prepare for interviews, and receive support to land paid internships in the field. **E**

Russakovsky codirects an AI4ALL summer program at Princeton, where high school students learn programming and AI basics and collaborate on research projects. They present their results at the end of the three-week session. *Photo by Lori Nichols*

MENGDI WANG MAKES A PLAY AT DECODING DISEASE

by Scott Lyon

On the windowsill of Mengdi Wang's Princeton office sits a box marked Civilization, a board game of strategy and conquest that starts in the Stone Age and ends with intergalactic flight.

"I love games," said Wang, an associate professor of electrical and computer engineering and an expert in artificial intelligence.

Wang spent a sabbatical year at Google DeepMind in 2020, studying techniques to improve reinforcement learning, the AI technology behind the world's most advanced Go-playing system, AlphaGo. While there, she noticed the beginnings of an explosive trend. From fusion energy to neuroscience, researchers of every stripe were turning to reinforcement learning and language models to solve hard, interesting problems.

Although language models were not her main specialty, she had studied them. In 2018, she wrote a paper with collaborators at Microsoft Research on reinforcement learning for natural language processing, the branch of machine learning that gave rise to today's large language models (think ChatGPT). But while she had been interested in the topic then, the rapid development of the field Wang saw in 2020 left her stunned. "The models had become so powerful," she said, "the set of solvable problems had become huge."

Two years later, generative AI thundered into public consciousness. And Wang turned her growing expertise in language models to advance another emerging technology, mRNA vaccines, like those that protect against COVID-19.

Working with the biotech startup RVAC Medicines, Wang developed a language model that treats biological sequences like text. Just as natural languages have syntax and grammar, the sequences of mRNA molecules dictate the forms and functions of proteins. And those proteins are the key ingredient in vaccine development.

Her model looks at a key region of mRNA molecules that modulates protein synthesis, generates variations in the genome sequences of that region, and identifies the most efficient, like rearranging words to create a more resonant sentence. The result, validated in wet-lab experiments, was a 33% improvement in protein-production efficiency — huge in this field.

Mengdi Wang. *Photo by Sameer A. Khan/Fotobuddy*

It's a foundational step toward what Wang called the life sciences' GPT moment, when AI breaks through to transform our relationship to health and disease. That's probably a long way off. It will require systematically unifying disparate branches of knowledge, a civilization-level feat. But it's there, over the horizon, like a perfect sentence waiting to be read. **E**

NEW JERSEY **AI HUB**

by Liz Fuller-Wright

Officials speaking at the AI hub announcement in December 2023 were (left to right) New Jersey Economic Development Authority CEO Tim Sullivan, Princeton Provost Jennifer Rexford, University President Christopher L. Eisgruber, New Jersey Gov. Phil Murphy, and Beth Noveck, the state's chief innovation officer. *Photo by Denise Applewhite*

New Jersey Gov. Phil Murphy and University President Christopher L. Eisgruber unveiled a plan last December to create an artificial intelligence innovation hub for the state in collaboration with the New Jersey Economic Development Authority.

The initiative will bring together AI researchers, industry leaders, startup companies, and other collaborators to advance research and development. Priorities include advancing the use of ethical AI and promoting workforce development in collaboration with other New Jersey universities, community colleges, and vocational schools.

"We have the potential to pioneer technologies that could unlock new cures for debilitating diseases, or new solutions for combating climate change, or new methods for educating our students so that every child can receive the personalized attention they deserve and need to reach their full potential," Murphy said. "With AI, we have a chance to confront — and perhaps overcome — some of the greatest challenges facing our world." **E**

PRECISION HEALTH INITIATIVE DEPLOYS DATA TO SURGICALLY TARGET MEDICAL CHALLENGES

by Julia Schwarz

Princeton researchers are working to make health care more precise and effective. But they're not working with patients, they're working with data.

Created in 2022, Princeton Precision Health is an interdisciplinary initiative for AI and data-driven approaches to health care challenges. The initiative uses advanced AI methods on large, complex data sets to make health care policy and delivery more precise, effective, and unbiased.

Led by Olga Troyanskaya, a professor of computer science and the Lewis-Sigler Institute for Integrative Genomics, PPH focuses on four areas: kidney disease and diabetes; the immune system and inflammation; neurodevelopment, neurodegeneration, and mental health; and technology and mental health.

The organization has hosted symposia, lectures, and discussion groups, and provided research grants to faculty. The initiative aims to convene experts from many fields — including computer science, biology, environmental science, social science, economics, policy, psychology, ethics, and medicine — to collaborate on interdisciplinary AI-based approaches to improve health outcomes.

"We can address very hard health challenges involving researchers across domains, from computational biology to epidemiology to social science and ethics," Troyanskaya said. **E**

Olga Troyanskaya, top right, leads a seminar for undergraduates on precision health research. *Photo by Tori Repp/Fotobuddy*

Barbara Graziosi, chair of the Department of Classics, described using large language models to analyze texts from Homer, Aristotle, and other ancient writers. *Photo by Sameer A. Khan/Fotobuddy*

NEW INITIATIVE **PUSHES BOUNDARIES** OF **LARGE LANGUAGE MODELS**

by Liz Fuller-Wright

The Princeton Language and Intelligence initiative is deploying the University's resources and academic strengths to enhance understanding of artificial intelligence; examine its safety, policy, and ethical implications; and enable its use across disciplines.

Announced in September 2023, the initiative is led by Sanjeev Arora, Princeton's Charles C. Fitzmorris Professor in Computer Science, who has conducted research at the intersection of language and AI for many years. The initiative will support a group of postdoctoral research fellows, research scientists, and engineers exploring key questions about the technology. It also has enabled the purchase a large computational cluster to give Princeton researchers access to the resources needed to work with large language models.

"AI is in danger of developing primarily inside private labs, with little opportunity for the rest of the world to know how it works or be assured that it is responsibly deployed with full safeguards," said Jennifer Rexford, Princeton University provost and Gordon Y.S. Wu Professor in Engineering. "We are committed to keeping AI expertise and know-how in the public sphere. We are eager to work with institutional partners and companies who share our commitment to open research. Initiatives such as PLI can democratize AI knowledge, enhance transparency and accountability, and help ensure that benefits of AI accrue to all of humanity." **E**

INITIATIVE TRAINS AI ON ENGINEERING

by John Sullivan

Reflecting rapid advances in artificial intelligence and machine learning, the School of Engineering and Applied Science's new AI Initiative will help accelerate innovation across engineering disciplines.

Led by Mengdi Wang, associate professor of electrical and computer engineering, and Ryan Adams, professor of computer science, the initiative, called AI Accelerated Innovation, aims to develop techniques to apply artificial intelligence to each step of technological innovation: design, simulation, fabrication, and control. The initiative builds on much work already underway at Princeton, such as using AI to design new energy technologies, medicines, computer chips, and construction techniques.

Such uses do not replace human intelligence, but extend it, Adams said. "How can we build systems that can do things for us that we cannot do for ourselves?"

The initiative will support new research through seed grants, postdoctoral positions, and computing resources, while providing internships, seminars, and other professional development.

The initiative will take advantage of Princeton's "tight-knit community of exceptionally high-quality researchers," the organizers said. "The goal is to create a collaborative, welcoming, resource-rich environment in which new research partnerships across disciplines can emerge and change the way engineering is done." E

Students delve into AI and its impacts.
With nearly half of computer science undergraduates studying artificial intelligence, demand for AI-related courses, independent work, and senior theses is booming. Out of 10 faculty members hired since 2022, half have expertise that adds to longstanding department strengths in computer vision, machine learning, natural language processing, and the ethical implications of AI. Teaching students to think about the social impact of AI, department chair Szymon Rusinkiewicz said, is as critical as teaching them how to build the technology. "Our goal is to make sure the computer science department at Princeton is well positioned to do both," he said. *Photos by Sameer A. Khan/Fotobuddy*

STATISTICS START TO UNTANGLE AI NETWORKS

by Steven Schultz

Boris Hanin, talking here with postdoctoral researchers Gage DeZoort and Mufan Li, is applying his background in mathematical physics to understand the structural properties of large neural networks that power AI systems. *Photo by Sameer A. Khan/Fotobuddy*

For AI systems to continue having transformative impacts in science and engineering, researchers may need to uncover fundamental rules that govern them, just as physicists did at key moments in the industrial revolution.

Boris Hanin, an assistant professor of operations research and financial engineering, is applying his background in mathematical physics — using math to develop theories about physical systems — to reveal a few simple rules that control neural networks, the massive algorithms that power much of today's artificial intelligence.

The birth of the modern steam engine in the late 1700s was "just excellent innovators doing inspired engineering, tinkering in labs," Hanin said. It wasn't until later in the 1800s that scientists fully laid out the laws of thermodynamics, the theoretical framework that led to entirely new technologies.

"You can't go from steam engine to jet engine without the notion of entropy," Hanin said.

Even though they are artificial, human constructions, giant AI models are like complex physical systems, Hanin said.

"It's actually quite possible to describe broadly the large-scale properties of neural networks in a similar way to how the Ideal Gas Law describes the large-scale properties of gases," Hanin said. In fact, the size and complexity of neural networks is what gives Hanin confidence. Scientists can, for example, easily say how turning up a thermostat will increase air pressure in a room even though they can't predict the individual behaviors of air molecules.

"That's the law of large numbers. It's the basis of all statistics," said Hanin.

The implications of such insights for AI systems could be enormous. Large neural networks generally perform better than smaller ones, but only after careful tuning. The larger the network, the more expensive it is to do this tuning by trial and error. A clear theoretical understanding of key variables that affect design choices could save a lot of time and money, Hanin said. "Part of the reason it's so costly to align large models is that we are still in the steam-engine phase of neural networks."

In recent research papers, Hanin and colleagues have begun to uncover meaningful relationships between the nature of neural networks and the statistics of the data they are designed to learn and interpret. "Little by little, this understanding is being extended," he said. E

IS **AI** TOO **POWERFUL** TO RELEASE **OPENLY?**

by Sayash Kapoor
and Arvind Narayanan

Arvind Narayanan and Matthew Salganik, professor of sociology, teach a course for graduate students on "Limits to Prediction." The course offers tools for critical examination of artificial intelligence and other prediction methods in realms ranging from geopolitics to weather to social media and the future of AI itself.
Photo by Sameer A. Khan/ Fotobuddy

Meta CEO Mark Zuckerberg said in January that his company plans to keep releasing and open-sourcing powerful AI.

The response was polarized. Some are excited by the potential for innovations that would be enabled when AI is openly available instead of being limited to those working at a big tech company. Others are alarmed, given that once AI is released openly, there's no stopping it from being used for malicious purposes, and call for policymakers to curb the release of open models.

The question of who should control AI development and who should have access to AI is of vital importance to society. Most leading models today (OpenAI's GPT-4, Anthropic's Claude 2, Google's Gemini) are closed: They can only be used via interfaces provided by the developers. But many others, such as Meta's Llama-2 and Mistral's Mixtral, are open: Anyone can download them, run them, and customize them. Their capabilities are a step below the leading closed models because of a disparity in resources used to train them. For example, GPT-4 reportedly cost over $100 million, whereas Llama-2 required under $5 million. This is another reason Zuckerberg's claims are interesting: He also announced that Meta is spending about $10 billion to acquire the computational resources needed to train AI. This means that the gulf in capabilities between open and closed models is likely to close or narrow.

Last fall, we collaborated with Stanford University to organize a virtual workshop to discuss the benefits and risks of openness in AI. We then assembled a team consisting of the organizers, many of the speakers, and a few collaborators to do an evidence review. What we found was surprising. Once we looked past the alarmism, we found very little evidence that openly released advanced AI, specifically large language models, could help bad actors more than closed ones (or even the non-AI tools available to them).

For example, a paper from MIT researchers claimed that AI could increase biosecurity risks. However, the information they found using open models was widely available on the internet, including on Wikipedia. In a follow-up study from RAND, a group of researchers in a simulated environment who had access to open models was no better at developing bioweapons than a control group that only had access to the internet. And when it comes to AI for cybersecurity, the evidence suggests that it helps defenders more than attackers.

We've noticed a pattern. Speculative, far-out risks such as bioterrorism, hacking nuclear infrastructure, or even AI killing off humanity generate headlines, make us fearful, and skew the policy debate. But the real dangers of AI are the harms that are already widespread. Training AI to avoid outputting toxic content relies on grueling work by low-paid human annotators who have to sift through problematic content, including hate speech and child sexual abuse material. AI has already led to labor displacement in professions such as translation, transcription, and creating art, after being trained using the creative work of these professionals without compensation. And lawyers have been sanctioned for including incorrect citations in legal briefs based on ChatGPT outputs, showing how overreliance on imperfect systems can go wrong.

Perhaps the biggest danger is the concentration of power in the hands of a few tech companies. Open models are in fact an antidote to this threat. They lower barriers to entry and promote competition. In addition, they have already enabled a vast amount of

Arvind Narayanan is a professor of computer science and director of Princeton's Center for Information Technology Policy (CITP). Sayash Kapoor is a computer science Ph.D. candidate at CITP whose research focuses on the societal impact of AI.

Narayanan and Kapoor are currently coauthoring a book, "AI Snake Oil," which looks critically at what AI can and cannot do.

CITP is an initiative of Princeton University's School of Engineering and Applied Science and the School of Public and International Affairs. The center's researchers work to better understand and improve the relationship between technology and society.

Photos of Sayash Kapoor and Arvind Narayanan by David Kelly Crow

Students discuss ideas with one another and with Professor Matthew Salganik (right) in the "Limits to Prediction" course. *Photo by Sameer A. Khan/Fotobuddy*

research on AI that could not be done without being able to download and examine the model's internals. They also benefit research that uses AI to study other scientific questions, say in chemistry or social science. For such research, they enable reproducibility. In contrast, closed model developers often deprecate or remove access to their older models, which leads to research based on these models being impossible to reproduce.

So far, we've mainly talked about language models. In contrast, for image or voice generators, open models have already been shown to pose significant risk compared to closed ones. Offshoots of Stable Diffusion, a popular open text-to-image model, have been widely used to generate non-consensual intimate imagery (NCII), including of real people. While Microsoft's closed model used to generate nude images of Taylor Swift was quickly fixed, such fixes are impossible to implement for open models — malicious users can remove guardrails because they have access to the model itself.

There are no easy solutions. Image generators have gotten extremely cheap to train and to run, which means there are too many potential malicious actors to police effectively. We think regulating the (mis)use of these models, such as cracking down on platforms used to distribute NCII, is much more justified and enforceable than regulating the development and release of the models themselves.

In short, we don't think policymakers should be rushing to put AI back in the bottle, although the reasons for our recommendation are slightly different for language models versus other kinds of generative AI models. While we should be cautious about the societal impact of large AI models and continue to regularly re-evaluate the risks, panic around their open release is unwarranted. An open approach will also make it easier for academia, startups, and hobbyists to contribute to building and understanding AI. It is vital to ensure that AI serves the public interest. We should not let its direction be dictated by the incentives of big tech companies. **E**

HOW DO YOU MAKE A ROBOT SMARTER?

by Molly Sharlach

Modern robots know how to sense their environment and respond to language, but what they don't know is often more important than what they do know.

Teaching robots to ask for help is key to making them safer and more efficient.

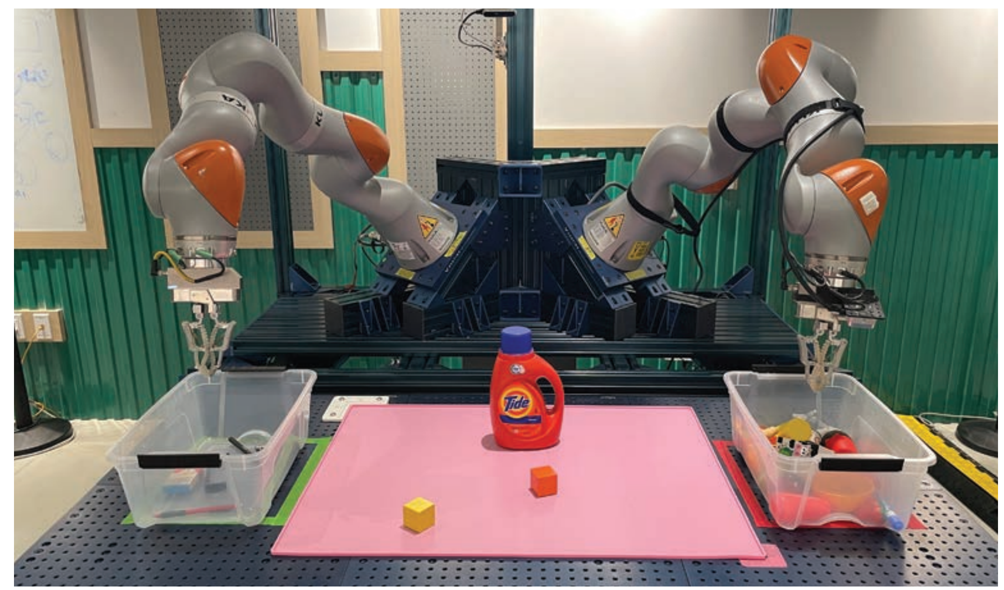
Engineers at Princeton University and Google have come up with a new way to teach robots to know when they don't know. The technique involves quantifying the fuzziness of human language and using that measurement to tell robots when to ask for further directions. Telling a robot to pick up a bowl from a table with only one bowl is fairly clear. But telling a robot to pick up a bowl when there are five bowls on the table generates a much higher degree of uncertainty — and triggers the robot to ask for clarification.

Because tasks are typically more complex than a simple "pick up a bowl" command, the engineers use large language models (LLMs) — the technology behind tools such as ChatGPT — to gauge uncertainty in complex environments. LLMs are bringing robots powerful capabilities to follow human language, but LLM outputs are still frequently unreliable, said Anirudha Majumdar, an assistant professor of mechanical and aerospace engineering at Princeton and the senior author of a study outlining the new method.

"Blindly following plans generated by an LLM could cause robots to act in an unsafe or untrustworthy manner, and so we need our LLM-based robots to know when they don't know," said Majumdar.

The system also allows a robot's user to set a target degree of success, which is tied to a particular uncertainty threshold that will lead a robot to ask for help. For example, a user would set a surgical robot to have a much lower error tolerance than a robot that's cleaning up a living room.

"We want the robot to ask for enough help such that we reach the level of success that the user wants. But meanwhile, we want to minimize the overall amount of help that the robot needs," said Allen Ren, a graduate student in mechanical and aerospace engineering at Princeton and the study's lead author. **E**



Engineers at Princeton University and Google developed a new way to teach robots to know when they don't know and ask for clarification from a human. The researchers tested their method on a simulated robotic arm and on two types of robot hardware. *Photo by the researchers*

THE NEXT AI FRONTIER? EXPANDING HARDWARE BY MAKING IT MORE COMPACT

by Scott Lyon

The physical facts of the AI boom are hard to fathom. Between 2012 and 2022, the computing power required by AI models grew by about 1 million percent, according to hardware expert Naveen Verma.

Cutting-edge digital chips boast circuit features at the scale of a virus, and still cannot keep pace with the accelerating computational demands of AI.

Naveen Verma. *Photo by Sameer A. Khan/Fotobuddy*

Verma, a professor of electrical and computer engineering, has developed a new kind of chip based on five key inventions from his lab. It promises a step change in AI hardware — 10 times more computational power per square millimeter and 10 times more energy efficiency than today's best chips.

The key lies in where and how all those trillions of mathematical operations are performed each second. To save time shuttling data between processor and memory, Verma found a way to compute directly in memory cells using analog signals, reimagining the physics of computation for modern workloads.

The Defense Advanced Research Projects Agency recently committed $18.2 million in funding to enable Verma and his colleagues to explore this technology's potential to power AI in laptops, phones, and vehicles. The project could lead to another massive expansion of AI's capabilities, especially in its broader use in safety-critical scenarios, from highways and hospitals to satellite orbits and beyond. **E**

THIS AI DOES MORE THAN THINK

by John Sullivan

Professor Ryan Adams and graduate student Cindy Zhang work with a laser cutter. *Photo by Tori Repp/Fotobuddy*

There are 27 bones in the human hand, but almost no one thinks about that when they pick up a water bottle. Ryan Adams does.

Adams, a professor of computer science, is leading research into mechanical intelligence. If artificial intelligence mimics cognition, the mechanical version acts like the autonomic system that drives walking, grasping, and breathing.

"Intelligence is more than just the brain," he said.

It is not only a matter of computation. Organisms' physical bodies are optimized for certain behaviors, which reduces the burden on the nervous system, said Adams.

"An example is grasping for primates. The design of your hand makes it easy to control," he said. "Fish bodies really want to swim. It makes it easy for the brain."

Over millions of years, these physical systems evolved in tandem with the neurological systems that control them. In Adams' lab, researchers deploy AI techniques to speed up this process. In one, they use a technique called gradient descent algorithms to allow AI systems to solve problems by trying many potential solutions. As the AI solves a single problem, it also develops the ability to solve other, similar problems.

"We can use things like gradients to get better at solving a problem," Adams said. "Can we use them to get better at mechanical things?" **E**

RESEARCHERS HARNESS LARGE LANGUAGE MODELS TO ACCELERATE MATERIALS DISCOVERY

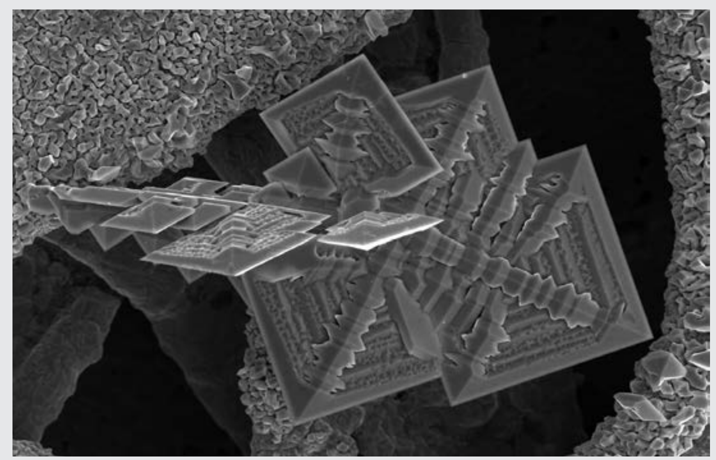
by Molly Sharlach

From salt to semiconductors, crystalline materials' capabilities are shaped not only by their atomic constituents, but also by the geometry of their microscopic structures.

But these attributes can be difficult to assess without making sample crystals in the lab. Now, a Princeton team led by Adji Bousso Dieng has created a method to evaluate crystals' capabilities without ever creating one. Instead of a materials lab, they are using artificial intelligence.

Dieng, an assistant professor of computer science, deploys large language models to synthesize information from descriptions of previously studied crystals, including details like the length and angles of bonds between atoms and measurements of electronic and optical properties. By leveraging decades of research, the new method could speed up the process of designing and testing new materials for technologies like batteries or semiconductors.

This approach "gives us a whole new way to look at the problem" of designing materials, said Craig Arnold, a co-researcher and a professor of mechanical and aerospace engineering. "It's really about, how do I access all of this knowledge that humanity has developed, and how do I process that knowledge to move forward?" **E**



A scanning electron microscope image of tin crystals, stimulated by electricity and growing on a copper surface. A new method developed by Princeton researchers could speed up the process of designing and testing new crystalline materials. *Image by Lynn Trahey, Argonne National Laboratory*

USING AI TO WRANGLE FUSION ENERGY

by Colton Poore

Research scholar Azarakhsh Jalalvand, Professor Egemen Kolemen, and PPPL postdoc Ricardo Shousha in the Andlinger Center lobby. *Photo by Adena Stevens*

Technicians stand inside a nuclear fusion experiment at the Princeton Plasma Physics Lab. The device, called a tokamak, is similar to the one at the DIII-D National Fusion Facility in California where Princeton Engineers tested their use of AI to control the super-heated plasmas that are critical to harnessing fusion as a clean source of energy. *Photo by Elle Starkman, PPPL Communications*

During a fusion reaction, every millisecond matters.

In the blink of an eye, the unruly, super-heated plasma that drives nuclear fusion can escape from its magnetic confinement within the donut-shaped device, or tokamak, designed to contain it. These getaways frequently spell the end of the reaction, posing a core challenge to developing fusion as a non-polluting, virtually limitless energy source.

But a team led by Egemen Kolemen *08, associate professor of mechanical and aerospace engineering and the Andlinger Center for Energy and the Environment, has trained an AI controller to predict and then avoid a type of plasma instability in real time.

The team first trained the controller on data from past experiments at the DIII-D National Fusion Facility in San Diego. They then demonstrated that the controller could learn from past experiments to predict the likelihood of instability during new fusion experiments and adjust specific reactor parameters in a matter of milliseconds to avoid the instability from ever forming.

The researchers showed their model could forecast potential plasma instabilities known as tearing mode instabilities up to 300 milliseconds in advance. While that leaves no more than enough time for a slow blink in humans, it was plenty of time for the AI controller to change certain operating parameters to avoid what would have developed into a tear within the plasma's magnetic field lines, upsetting its equilibrium and opening the door for a reaction-ending escape.

"By learning from past experiments, rather than incorporating information from physics-based models, the AI could develop a final control policy that supported a stable, high-powered plasma regime in real time, at a real reactor," said Kolemen, who is also a staff research physicist at the Princeton Plasma Physics Laboratory (PPPL).

The research opens the door for more dynamic control of a fusion reaction than current approaches, and it provides a foundation for using artificial intelligence to solve a broad range of plasma instabilities, which have long been obstacles to achieving a sustained fusion reaction.

"Previous studies have generally focused on either suppressing or mitigating the effects of these tearing instabilities after they occur in the plasma," said first author Jaemin Seo, an assistant professor of physics at Chung-Ang University in South Korea who performed much of the work while a postdoctoral researcher in Kolemen's group. "But our approach allows us to predict and avoid those instabilities before they ever appear."

While the researchers said the work is a promising proof of concept demonstrating how artificial intelligence can effectively control fusion reactions, it is only one of many next steps already ongoing in Kolemen's group to advance the field of fusion research.

The first step is to get more evidence of the AI controller in action at the DIII-D tokamak, and then expand the controller to function at other tokamaks.

A second line of research involves expanding the algorithm to handle many different control problems at the same time. While the current model uses a limited number of diagnostics to avoid one specific type of instability, the researchers could provide data on other types of instabilities and give access to more knobs for the AI controller to tune, potentially allowing it to control for several types of instabilities simultaneously.

And on the route to developing better AI controllers for fusion reactions, researchers might also gain more understanding of the underlying physics. By studying the AI controller's decisions as it attempts to contain the plasma, which can be radically different than what traditional approaches might prescribe, artificial intelligence may be not only a tool to control fusion reactions but also a teaching resource.

"Eventually, it may be more than just a one-way interaction of scientists developing and deploying these AI models," said Kolemen. "By studying them in more detail, they may have certain things that they can teach us too." **E**

PRINCETON UNIVERSITY

School of Engineering and Applied Science
Princeton, NJ 08544-5263
www.princeton.edu/engineering

eqn@princeton.edu