

Spring 2025

AI: PROMISES AND RISKS

The

BRIDGE

LINKING ENGINEERING AND SOCIETY

Toward an Evaluation Science for Generative AI Systems

Laura Weidinger, Deb Raji, Hanna Wallach, Margaret Mitchell, Angelina Wang, Olawale Salaudeen, Rishi Bommasani, Sanmi Koyejo, and William Isaac

Dashboards for AI: Models of the User, System, and World

Fernanda Viégas and Martin Wattenberg

The Next Frontier in AI: Understanding the 3-D World

Fei-Fei Li

AI's Capabilities Make It a Powerful Tool for Driving Societal Impact

Yossi Matias, Avinatan Hassidim, and Philip Nelson

Disrupting the Disruption Narrative: Policy Innovation in AI Governance

Alondra Nelson

Alignment of AI Systems' Risk Attitudes, and Four Real-Life Examples

Elisabeth Paté-Cornell

Unlocking the Gut's Brain with Ingestible Bioelectronics

Shriya Srinivasan

NATIONAL ACADEMY OF ENGINEERING

The mission of the National Academy of Engineering is to advance the welfare and prosperity of the nation by providing independent advice on matters involving engineering and technology, and by promoting a vibrant engineering profession and public appreciation of engineering.

The BRIDGE

NATIONAL ACADEMY OF ENGINEERING

Erroll B. Davis Jr., *Chair*
John L. Anderson, *President*
Wesley L. Harris, *Vice President*
Howard B. Rosen, *Home Secretary*
Nadine Aubry, *George and Virginia Bugliarello International Secretary*
Roger L. McCarthy, *Treasurer*

Editor in Chief: Ronald M. Latanision
Editor: Kyle Gipson
Production Associate: Penelope Gibbs

The Bridge (ISSN 0737-6278) is published quarterly by the National Academy of Engineering, 500 Fifth Street NW, Washington, DC 20001. Periodicals postage paid at Washington, DC.

Vol. 55, No. 1, Spring 2025

Postmaster: Send address changes to *The Bridge*, 500 Fifth Street NW, Washington, DC 20001.

Changes of address or requests to unsubscribe should be sent to PGibbs@nae.edu.

Papers are presented in *The Bridge* on the basis of general interest and timeliness. They reflect the views of the authors and not necessarily the position of the National Academy of Engineering.

The Bridge is printed on recycled paper. ♻️

© 2025 by the National Academy of Sciences. All rights reserved.

Mission Statement of *The Bridge*

The Bridge publishes articles on engineering research, education, and practice; science and technology policy; and the interface between engineering and technology and society. The intent is to stimulate debate and dialogue both among members of the National Academy of Engineering (NAE) and in the broader community of policymakers, educators, business leaders, and other interested individuals. *The Bridge* relies on its editor in chief, NAE members, and staff to identify potential issue topics and guest editors. Invited guest editors, who have expertise in a given issue's theme, are asked to select authors and topics, and independent experts are enlisted to assess articles for publication. The quarterly has a distribution of about 7000, including NAE members, members of Congress, agency officials, engineering deans, department heads, and faculty, and interested individuals all over the country and the world. Issues are freely accessible at www.nae.edu/TheBridge.

A complete copy of *The Bridge* is available in PDF format at www.nae.edu/TheBridge. Some of the articles in this issue are also available as HTML documents and may contain links to related sources of information, multimedia files, or other content.

The

Volume 55, Number 1 • Spring 2025

BRIDGE

LINKING ENGINEERING AND SOCIETY

A Word from the NAE Chair

3 Expanding the US Engineering Workforce

Erroll B. Davis Jr.

5 Editor in Chief's Note

Ronald M. Latanision

Guest Editors' Note

7 Realizing the Transformative Potential of AI

William Isaac and Marian Croak

Features

9 Toward an Evaluation Science for Generative AI Systems

Laura Weidinger, Deb Raji, Hanna Wallach, Margaret Mitchell, Angelina Wang, Olawale Salaudeen, Rishi Bommasani, Sanmi Koyejo, and William Isaac

There is an urgent need for a more robust and comprehensive approach to AI evaluation.

19 Dashboards for AI: Models of the User, System, and World

Fernanda Viégas and Martin Wattenberg

AI dashboards can promote transparency and user trust, enabling more effective human-AI interaction.

27 The Next Frontier in AI: Understanding the 3-D World

Fei-Fei Li

AI is being developed to understand and interact with the 3-D world, opening up new possibilities in fields like robotics and healthcare.

33 AI's Capabilities Make It a Powerful Tool for Driving Societal Impact

Yossi Matias, Avinatan Hassidim, and Philip Nelson

AI innovations, developed and deployed responsibly, can help preserve our climate, improve health outcomes, and create a more accessible world for everyone.

42 Disrupting the Disruption Narrative: Policy Innovation in AI Governance

Alondra Nelson

Governance should not be understood as an impediment to AI innovation but as an essential component of it.

49	Alignment of AI Systems’ Risk Attitudes, and Four Real-Life Examples <i>Elisabeth Paté-Cornell</i> AI systems’ risk preferences should be consistent with those of humans, particularly in critical domains like healthcare and national security.
57	Unlocking the Gut’s Brain with Ingestible Bioelectronics <i>Shriya Srinivasan</i> Ingestible devices present a novel approach to accessing the gastrointestinal tract, making possible therapeutic interventions that can modulate gut function and potentially improve health outcomes.
61	Op-Ed: Why Engineers Should Learn Economics <i>Debasis Mitra</i> Engineers and economists have much to learn from each other, and there is much to be gained from them working collaboratively.
<hr/>	
	News and Notes
65	NAE Members Awarded National Medals of Science and of Technology and Innovation
67	Class of 2025 Elected
75	NAE Newsmakers
78	2025 Fritz J. and Dolores H. Russ Prize Acceptance Remarks by Ian Shanks
79	Engineering Innovation for a Brighter Future: The 2025 NAE National Meeting
81	Message from NAE Vice President Wesley L. Harris
83	2024 Honor Roll of Donors
95	Calendar of Meetings and Events
96	In Memoriam
98	<i>Invisible Bridges: Ports, Politics, and Peace:</i> The Engineering of Stability

NATIONAL ACADEMIES Sciences Engineering Medicine

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected

by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at www.nationalacademies.org.

A Word from the NAE Chair

Expanding the US Engineering Workforce



Erroll B. Davis Jr. (NAE), chair, the National Academy of Engineering

As some would note, I am fond, perhaps too fond, of saying, “May you live in interesting times.” I have always lived in “interesting times,” ranging from working inside a UNIVAC I 8K memory unit in the early 1960s to working as the “computer jock” for Nobel Laureate Herb Simon, one of the early pioneers of artificial intelligence (AI), to serving as an Army officer during the Vietnam War to embarking on careers in business and education, in both higher ed and K-12. Obviously, no one would dispute that we are very much living in “interesting times” today. We are living in a time of political intensity with rapidly developing AI technology (and its significant impacts). Both forces suggest that we need more engineers from diverse backgrounds if we are to maintain or increase our leadership in an evolving technological landscape. Let me offer a personal example of why we need more people from diverse backgrounds in the engineering workforce as we venture further into the exciting world of AI.

My wife and I often go grocery shopping together on the weekends. On one particular weekend, we decided to visit a recently remodeled and modernized supermarket, which is part of a popular chain. After we entered the store, we both had to use the restroom. Before exiting, I washed my hands as usual. My wife had a different experience: As she emerged from the restroom, she informed me that the automatic water dispensers in the sinks seemed to be working for others but not for her! I thought for a moment, and then I suggested that she run an experiment. I suggested that she take a piece of white

toilet paper in one hand as she waved it by the sensor to see if she got a different outcome. She did. The automatic water dispenser worked when it detected a lighter skin tone. Her hands are a bit darker than mine, and I did not encounter the problem that she did.

Obviously, this was not the end of the world for either of us, but it did recall several lectures I attended given by Carnegie Mellon professors on the inherent biases contained in AI algorithms. Did someone deliberately program skin tone biases into the sink water dispenser? I think not. But I also believe that if a darker-skinned engineer had been on the design team, my wife would have had a better outcome. AI is but one engineering field where we need more engineers and more diverse engineers. As AI continues to develop, I believe that we must have a diverse engineering workforce that reflects a broad cross section of the US population.

This same point about developing more homegrown engineers from diverse backgrounds is echoed in the ongoing intense political debate about allowing foreign-born engineers to work on visas in Silicon Valley and elsewhere in America. One side, of course, wants to continue to import well-trained and highly skilled engineers. The other side wants to immediately restrict immigration, but that side is also calling for investments in education and the production of more US-born or naturalized engineers. From my perspective, this is a positive because once you go down the path of cultivating more engineers, if you are serious about it, you will need to

reach into more diverse populations in order to achieve any meaningful outcomes. I firmly believe that this will benefit the engineering sector and America more broadly as we incorporate more perspectives into our engineering work, particularly in AI engineering.

We need more well-trained and qualified engineers in this country, and we are not going to get the numbers we need from a small sector of our society. We will need engineers from every sector if America is to remain globally competitive. Sheer demand and changing demographics reinforce this sentiment. In “The Uncapped Potential: Engineering an Opportunity of a Lifetime,”¹ which appeared in the fall 2024 issue of *The Bridge*, Norm

Augustine makes this very point, drawing on a plethora of data. If we can increase the representation of women and other currently underrepresented groups in the engineering profession, we will not just get a few more engineers. We will get hundreds of thousands of them! I encourage you to read the article. Augustine makes it clear that it is time to do the work at both the national and grassroots levels. It is a step in the right direction for the engineering profession to recognize the problem of the future engineering workforce. However, it should go beyond recognizing the problem; the engineering profession should take some responsibility for solving it.

¹ Augustine NR. 2024. The uncapped potential: Engineering an opportunity of a lifetime. *The Bridge* 54(3):10–18.

Editor in Chief's Note



Ronald M. Latanision (NAE) is a senior fellow at Exponent, the Neil Armstrong Distinguished Visiting Professor at Purdue University, and editor in chief of *The Bridge*.

First, I want to thank William Isaac and **Marian Croak** for serving as guest editors of this issue on artificial intelligence (AI). They have assembled an issue that touches all the bases in terms of the pressing matters regarding AI, ranging from visual instrumentation, track system conditions, and performance standards and risks to governance, regulatory guardrails, and, ultimately, social impact.

There is a lot of good that has come from AI—in health care and the development of new materials, for example. And there are surely more positive developments to come. But I am equally certain that this technology, just as others, can be and is being used abusively. The changes in AI that began in November 2022 with the release of generative AI are truly remarkable. GenAI is not just another new technology; it has the potential to revolutionize the way we work and live. The development of GenAI is earth-shattering. One could say that, to the average thoughtful person, the introduction of the telephone or the Ford Model T must have been just as momentous. What is different in the case of GenAI is that it does not just add a new dimension to our lives; it presents technology as a force beyond nature. GenAI apparently thinks and feels, though it is not yet clear on what scale and in what detail relative to human thinking. Granted, we don't really understand the particulars of how humans think either.

The projected proliferation of AI attendant on the recent DeepSeek announcements, if fully realized, would make AI even more daunting and essentially unencumbered by fiscal concerns. I worry that this technology may be heading so far out in front of humans that people

may begin to broadly distrust science and technology on a level that is unprecedented. That erosion of trust would be to our collective misfortune.

Technology and technologists have crucial roles to play in advancing medicine, meeting energy demand, addressing climate change, improving K-12 education, and so much more. Prior to the recent advances in AI, I hoped that we had learned useful lessons from the history of the internet and the web that would lead to a responsible and accountable integration of AI into our social fabric. But I do not see evidence that suggests that we have learned much of anything from this history. My sense, however, is that GenAI has the potential to be supremely useful and also supremely abusive (personally, socially, and culturally). We must all be concerned about reducing risks and ensuring that GenAI is used in constructive and societally beneficial ways.

Like any technology, the future of GenAI will be determined by how people choose to use it: for good purposes or bad. I am confident that it will be used for both. That is why it is so essential that we introduce GenAI in ways that maximize its potential benefits and anticipates and reduces potential harms, for individuals and for society.

Technologists design engineering systems based on verifiable facts. To do otherwise would lead to the failure of such systems. And the same applies in our contemporary culture in many ways. Who would, for example, trust a surgeon to operate without valid, fact-based diagnostics? Technologists must look to facts for validation in designing engineering systems that work. Could AI be trained to solve its own problems? For example, could it be required to train on validated data?

We should strive to ensure that generative AI systems are based on reliable facts and evidence-based research. We—scientists, engineers, and technologists—must ensure that our work is grounded in quality data and truth. AI developers should be required to adhere to these same standards. At the same time, our goals must be the responsible development and introduction of generative AI. We should work toward a society of bots and humans that co-exist on terms that preserve rather than destroy the best that humans have to offer.

Finally, I want to acknowledge two gentlemen with whom I share many conversations about technology in general but particularly about AI: Ron Smith of Innova-

tion Toronto and Marv Goldschmitt of Bedford, Massachusetts. Ron and Marv both have distinguished careers at the leading edge of technology, and they have added a freshness to my thinking that I treasure. We don't always agree, but we are never personally disagreeable.

For this issue, we planned to include an interview, which we conducted in December 2024, with an engineer who worked to address public policy matters. Given the rapid shifts that have accompanied the new presidential administration, that interviewee requested that we not publish the interview. We honored that request.

As always, I welcome your comments. Feel free to reach out to me at rlatanision@alum.mit.edu.

Guest Editors' Note

Realizing the Transformative Potential of AI



William Isaac



Marian Croak

William Isaac is a principal scientist and head of responsible research at Google DeepMind. Marian Croak (NAE) is vice president of Society-Centered AI and Foundational ML at Google.

The rapid rise of artificial intelligence (AI) presents a defining moment in human history. The past year bore witness to both exhilarating progress and growing anxieties surrounding AI's expanding role in society. The advent of generative AI, capable of crafting human-quality text and visuals, has ignited widespread innovation. Yet, this progress has also spurred concerns about value alignment, safety, and misuse and misinformation. Across the globe, regulatory and geopolitical concerns are emerging as the technology becomes more capable and competition increases. As AI becomes increasingly intertwined with our lives, the urgent need to address these challenges and ensure responsible AI development and deployment, guided by human values and societal well-being, has become paramount.

This issue of *The Bridge* delves deep into this complex landscape, offering a rich tapestry of perspectives on AI's promises and challenges. Articles in this issue explore critical themes such as AI evaluation science, the imperative of transparency and user trust, the emergence of spatial intelligence, AI's potential in tackling societal challenges, policy innovation in AI governance, and the alignment of AI systems and human attitudes towards risk.

Several cross-cutting themes emerge from these contributions, underscoring the interconnected nature of AI's various facets. One prominent theme is the crucial importance of responsible AI development and deployment, encompassing safety, reliability, and alignment with

human values. Another key theme is the need for transparency and user trust, enabling users to comprehend AI systems' workings and limitations as they become more sophisticated and integrated into our lives. Additionally, the articles in this issue emphasize the transformative potential of AI across diverse domains, from healthcare and education to robotics and governance, while recognizing the necessity for careful consideration of ethical and societal implications.

In this issue:

- Laura Weidinger, Deb Raji, Hanna Wallach, Margaret Mitchell, Angelina Wang, Olawale Salaudeen, Rishi Bommasani, Sanmi Koyejo, and William Isaac illuminate the urgent need for a more robust and comprehensive approach to AI evaluation in "Toward an Evaluation Science for Generative AI Systems." They propose an evaluation science for AI, drawing lessons from other fields such as medicine and civil engineering, where evaluation has played a critical role in ensuring safety and reliability.
- Fernanda Viégas and Martin Wattenberg explore the concept of AI dashboards as a way to provide real-time information about the internal states of AI systems in "Dashboards for AI: Models of the User, System, and World." They argue that such dashboards can promote transparency and user trust, enabling more effective human-AI interaction.

- **Fei-Fei Li** discusses the exciting advancements in computer vision and spatial intelligence in “The Next Frontier in AI: Understanding the 3-D World.” Li highlights how AI is being developed to understand and interact with the 3-D world, opening up new possibilities in fields like robotics and healthcare.
- **Yossi Matias, Avinatan Hassidim, and Philip Nelson** provide compelling examples of AI innovations that are helping to preserve our climate, improve health outcomes, and create a more accessible world for everyone in “AI’s Capabilities Make It a Powerful Tool for Driving Societal Impact.” The authors emphasize the importance of responsible AI development and deployment to ensure that these benefits are realized for all.
- **Alondra Nelson** challenges the prevailing notion that AI innovation outpaces policy development in “Disrupting the Disruption Narrative: Policy Innovation in AI Governance.” Nelson advocates for proactive and innovative policymaking to ensure that AI technologies are developed and used responsibly.
- **Elisabeth Paté-Cornell** examines the alignment of AI systems’ risks with those of human decision-makers in “Alignment of AI Systems’ Risk Attitudes, and Four

Real-Life Examples.” Paté-Cornell discusses the importance of ensuring that AI systems’ risk preferences are consistent with those of humans, particularly in critical domains like healthcare and national security.

As AI continues its relentless advance, we must confront critical questions about its future. How can we guarantee that AI benefits all of humanity? What are the ethical and societal implications of increasingly sophisticated AI systems? How can we foster transparency and user trust in AI? And how can we govern AI in a way that promotes innovation while safeguarding against potential risks?

These are just a few of the questions that demand our attention as we navigate the transformative landscape of AI. The articles in this issue offer valuable insights and perspectives on these and other crucial issues. We hope that they will stimulate further discussion and debate, ultimately contributing to a more informed and responsible approach to AI development and deployment.

We express our sincere gratitude to all the authors for their insightful contributions to this special issue. We also extend our appreciation to the entire *Bridge* editorial team for their tireless efforts in bringing this issue to fruition. We trust that you will find these articles both informative and thought-provoking.

There is an urgent need for a more robust and comprehensive approach to AI evaluation.

Toward an Evaluation Science for Generative AI Systems

Laura Weidinger, Deb Raji, Hanna Wallach, Margaret Mitchell, Angelina Wang, Olawale Salaudeen, Rishi Bommasani, Sanmi Koyejo, and William Isaac



Laura Weidinger



Deb Raji



Hanna Wallach



Margaret Mitchell



Angelina Wang



Olawale Salaudeen



Rishi Bommasani



Sanmi Koyejo



William Isaac

There is an increasing imperative to anticipate and understand the performance and safety of generative artificial intelligence (AI) systems in real-world deployment contexts. However, the current evaluation ecosystem is insufficient:

Laura Weidinger* is a staff research scientist at Google DeepMind. Deb Raji* is a doctoral candidate at the University of California, Berkeley. Hanna Wallach is a VP and distinguished scientist at Microsoft Research. Margaret Mitchell is a researcher and chief ethics scientist at Hugging Face. Angelina Wang is a postdoc at Human-Centered Artificial Intelligence and RegLab, Stanford University. Olawale Salaudeen is a postdoctoral associate in the Laboratory for Information and Decision Systems at the Schwarzman College of Computing, Massachusetts Institute of Technology. Rishi Bommasani is the society lead at the Stanford Center for Research on Foundation Models. Sanmi Koyejo is an assistant professor of computer science at Stanford University. William Isaac is a principal scientist and head of responsible research at Google DeepMind.

* contributed equally to this work

Commonly used static benchmarks face validity challenges, and ad hoc case-by-case audits rarely scale. In this piece, we advocate for maturing an evaluation science for generative AI systems. While generative AI creates unique challenges for system safety engineering and measurement science, the field can draw valuable insights from the development of safety evaluation practices in other fields, including transportation, aerospace, and pharmaceutical engineering. In particular, we present three key lessons: Evaluation metrics must be applicable to real-world performance, metrics must be iteratively refined, and evaluation institutions and norms must be established. Applying these insights, we outline a concrete path toward a more rigorous approach for evaluating generative AI systems.

The Rise of Generative AI Systems

The widespread deployment of generative AI systems in medicine (Boyd 2023), law (e.g., Lexis+AI¹), education (Singer 2024), information technology (e.g., Microsoft's Copilot²; Reid 2024), and many social settings (e.g., Replika,³ character.ai) has led to a collective realization: The performance and safety of generative AI systems in real-world deployment contexts are very often poorly anticipated and understood (Mulligan 2024; Roose 2024a; Wiggers 2024). The tendency of these systems to generate inaccurate statements has already led to the spread of medical and other misinformation (Archer and Elliott 2025; Omiye et al. 2023); incorrect legal references (Magesh et al. 2024); failures as educational support tools (Singer 2023); and widespread confusion in search engine use (Heaven 2022; Murphy Kelly 2023). Beyond factual discrepancies, AI-enabled chatbots have also been described as interacting inappropriately with users (Roose 2023), exposing security vulnerabilities (Nicolson 2023), and fostering unhealthy emotional reliance (Dzieza 2024; Roose 2024b; Verma 2023).

The historical focus on benchmarks and leaderboards has been effective at encouraging the AI research community to pursue shared directions; however, as AI products become widely integrated into our everyday lives, it is increasingly clear that static benchmarks are not well suited to improving our understanding of the real-world performance and safety of deployed generative AI systems (Bowman and Dahl 2021; de Vries et al. 2020; Goldfarb-Tarrant et al. 2021; Liao et al. 2021; Raji 2021). Despite this mismatch, static benchmarks are still

commonly used to inform real-world decisions about generative AI systems that stretch far beyond the research landscape—such as in deployment criteria and marketing materials for new model or system releases (Anthropic 2024b; Gemini Team Google 2024; Grattafiori et al. 2024; OpenAI 2024), third party critiques (Mirzadeh et al. 2024; Zhang et al. 2024), procurement guidelines (Johnson et al. 2024), and in public policy discourse (NIST 2023; European Commission 2024). Although there is an emerging interest in more interactive (Chiang et al. 2024), dynamic (Kiela et al. 2021), and behavioral (Ribeiro et al. 2020) approaches to evaluation, many of the existing alternatives to benchmarks, such as red teaming exercises and case-by-case audits, still fall woefully short of enabling systematic assessments and accountability (Birhane et al. 2024; Friedler et al. 2023).

For AI evaluation to mature into a proper “science,” it must meet certain criteria. Sciences are marked by having theories about their targets of study, which can be expressed as testable hypotheses. Measurement instruments to test these hypotheses must provide experimental consistency (i.e., reliability, internal validity) and generalizability (i.e., external validity). Finally, sciences are marked by iteration: Over time, measurement approaches and instruments are refined and new insights are uncovered. Collectively, these properties of sciences contrast sharply with the practice of rapidly developing static benchmarks for evaluating generative AI systems, while anticipating that within a few months such benchmarks will become much less useful or obsolete.

As generative AI exits an era of research and enters a period of widespread use (Hu 2023; Reid 2024), the field risks exacerbating an ongoing public crisis of confidence in AI technology (Faveiro and Tyson 2023) if we do not develop a more mature evaluation science for generative AI systems. From the history of other fields, we can get a sense of why: Collectively, leaderboards, benchmarks, and audits do not amount to the robust and meaningful evaluation ecosystem we need to properly assess the suitability of these products in widespread use. In particular, they cannot give assurances about AI system performance in different domains or for different user groups.⁴ In this piece, we advocate for the maturation of such an evaluation science. By drawing on insights from systems safety engineering and measurement science in other fields,

⁴ We thank our anonymous reviewer for pointing out that one reason for this is that benchmarks are often not robust to data or domain shifts (i.e., benchmarks test AI system outputs in certain contexts but this may not be predictive of AI system behaviour in other contexts).

¹ <https://www.lexisnexis.com/en-us/products/lexis-plus-ai.page>

² copilot.microsoft.com

³ replika.com

while acknowledging the unique challenges inherent to generative AI, we identify three important properties of any evaluation science that the AI community will need to focus on to meaningfully advance progress: a focus on real-world applicability, iterative measurement refinement, and adequate institutional investment. These properties then enable us to outline a concrete path toward a more rigorous evaluation ecosystem for generative AI systems.

Lessons from Other Fields

The bridges we stand on, the medicine we take, and the food we eat are all the result of rigorous assessment. In fact, it is because of the rigor of the corresponding evaluation ecosystems that we can trust that the products and critical infrastructure surrounding us are performant and safe. Generative AI products are no exception to this reality and therefore not unique in their need for robust evaluations. In response to their own crises, more established evaluation regimes emerged in other fields to assure users and regulators of safety and reliability—offering concrete lessons for the AI field (Raji and Dobbe 2023; Raji, Kumar et al. 2022; Rismani et al. 2023). We note three key evaluation lessons from these other fields: the targeting of real-world performance, the iterative refinement of measurement approaches, and the establishment of functioning processes and institutions.

Real-World Applicability of Metrics

First, it is noteworthy that, historically, evaluation made a difference for safety because it tracked real-world risks. Measuring real-world performance does not mean waiting until risks manifest—on the contrary, earlier pre-deployment risk detection and evaluation allows for more comprehensive and cheaper mitigations (Collingridge 1982). For example, in clinical trials, strict requirements exist for staged, pre-clinical testing in order to minimize risks to vulnerable patient populations. Similarly, airplanes are first designed and tested through simulations to improve understanding of their performance while minimizing risks to life and material damage.

Pre-deployment testing may help identify real-world risks earlier—however, it must be accompanied by post-deployment monitoring to detect emergent harms as they happen. For instance, unexpected side effects and off-label use of pharmaceuticals in the medical domain, especially on under-tested populations, are nearly impossible to anticipate pre-deployment. Many of these issues only emerge from highly complex interactions at the point of use. In such cases, health providers, patients, and manu-

facturers are required to report adverse events to regulatory agencies via incident databases.⁵ The collection of these incidents and the resulting analyses can then be used to inform any restriction or cautionary uses of the drug or vaccine, especially for at-risk populations. As an example, the discovery of myocarditis symptoms from the COVID-19 vaccine was facilitated by the Vaccine Adverse Events System (VAERS) incident database. This finding led to a warning and an adjusted dosage recommendation for the most impacted population of male vaccine recipients, aged 12 to 17 (Oster et al. 2022). In some cases, monitoring data can even be used to feed back into future pre-deployment evaluation practices—for example, the results of race-based failures observed in an FDA incident database for medical devices (FDA MAUDE Database⁶) informed new health department guidelines on adequate equitable representation in pre-clinical trials for such devices (Fox-Rawlings et al. 2018; US Food & Drug Administration 2017).

As generative AI exits an era of research and enters a period of widespread use, the field risks exacerbating an ongoing public crisis of confidence in AI technology if we do not develop a more mature evaluation science for generative AI systems.

Iteratively Refining Metrics

The metrics and measurement approaches of evaluation must be iteratively refined and calibrated over time. This iterative process includes choosing and refining relevant measurement targets (i.e., the concepts to be measured). Initially, the automotive industry focused on human-

⁵ For example: <https://open.fda.gov/data/faers/>, <https://vaers.hhs.gov/>, <https://yellowcard.mhra.gov.uk/>.

⁶ www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm

caused errors, responding with drivers' education, drivers' licenses, and laws against drunk driving. However, as accidents continued to soar, seatbelt regulations and other design choices became a focal point, feeding into notions of a car's "crashworthiness" tied to manufacturer responsibility (Díaz and Costas 2020). This measurement target of crashworthiness has continued to evolve over time. For example, in Europe, concerns about the safety of pedestrians and cyclists were incorporated in an expanded notion of crashworthiness (UN 2011), broadening what it means for a car to be considered "safe."

As a measurement target is refined over time, so are the measurement instruments that are designed to capture it. With the measurement of temperature, divergent thermometer readings revealed the importance of engineering instruments with a reliable liquid indicator (Chang 2001). Further attempts to calibrate thermometers gave rise to deeper insights about temperature itself—as an indication of matter phase changes (i.e., Celsius), human body responses (i.e., Fahrenheit), and quantum mechanical properties (i.e., Kelvin). However, no single measurement instrument is perfect—by triangulating results from multiple methods, more robust insights can be gained (Campbell and Fiske 1959; Jespersen and Wallace 2017). Ultimately, identifying measurement targets, designing metrics, and developing measurement instruments are all interdependent tasks that require a careful iterative process.

Establishing Institutions & Norms

A successful evaluation ecosystem requires investing in institutions. The advocacy of Harvey Wiley, Samuel Hopkins Adams, and others led to the 1938 passing of the United States Federal Food, Drug, and Cosmetic Act. This act led to the creation of the Food and Drug Administration (FDA), an agency that is now widely known for its rigorous pharmaceutical and nutrition testing regimes. At the FDA, Wiley and his team developed numerous innovative methods for identifying the presence and effects of particular poisonous ingredients, notably leading several multi-year experiments to assess the pernicious effects of various chemicals on a group of volunteers known as the "Poison Squad" (Blum 2018). Without the centralization of testing efforts through a single agency, this team could not have had the resources or coordination capacities to execute such long-term and large-scale experiments.

In many fields, readily available evaluation tools, shared evaluation infrastructure, and standards afforded by such

institutions have contributed to the establishment of more thorough evaluation regimes (Timmermans and Berg 2003; Vedung 2017). After the number of cars on the road increased by an order of magnitude throughout the early 20th century, the corresponding increase in fatal crashes pushed Ralph Nader and other advocates to establish the National Traffic and Motor Vehicle Safety Act in 1966, responsible for the National Highway Safety Bureau (now the transportation testing agency known as the National Highway Traffic Safety Administration). By 1985, Ralph Nader claimed, "Programs, which emphasize engineering safety, have saved more than 150,000 lives and prevented or reduced in severity a far larger number of injuries" (Nader 1985). In 2015, an NHTSA report revealed that this trend has continued, with an estimated 613,501 lives saved between 1960 and 2012 (Britannica 2025). Nader attributed much of this success to the meaningful enforcement of government-mandated standards, including active monitoring (i.e., regularly measuring everything from fuel efficiency to auto handling and braking capabilities) by the National Highway Traffic Safety Agency, which led to the recall of millions of defective vehicles and tires by the early 1980s.

Towards an Evaluation Science for Generative AI

Unique Challenges of Generative AI

While drawing on lessons from other fields, it is important to understand what makes the challenge of evaluating generative AI systems unique. Other systems—from personal computers to pharmaceuticals—can be used for purposes that were not originally intended. However, generative AI systems are often explicitly designed to be open-ended—that is, underspecified and deliberately versatile in the range of use cases they support (Hughes et al. 2024). This open-endedness makes it hard to define precise measurement targets in AI evaluation, resulting in vague targets such as the long-standing trend of measuring an AI system's "general intelligence," rather than performance on specific tasks (Raji et al. 2021). Furthermore, generative AI systems tend to be less deterministic—the same input can lead to different outputs due to their stochastic nature and due to unknown factors in training data (Raji 2021). This non-determinism makes it harder to predict system behaviors compared to prior software systems, as it is difficult to directly trace system design choices—about training data, model design, or the user interface—to downstream system outputs and impacts.

Further adding to the complexity of anticipating and evaluating AI system outputs and use cases is the pos-

sibility of longitudinal social interactions with generative AI systems. This gives rise to a new class of interaction risks that may evolve in unexpected ways over time (e.g., harmful human–AI “relationships” [Manzini et al. 2024]). Taken together, these unique challenges inherent to generative AI systems indicate the need for a behavioral approach to evaluating such systems, focusing on AI system performance in the context of different real-world settings (Matias 2023; Rahwan et al. 2019; Wagner et al. 2021). Indeed, adopting a behavioral approach that treats AI systems as black boxes can be helpful in enabling some translation between higher-level systemic impact evaluations and lower-level computational methods (McCoy et al. 2024; Shiffrin and Mitchell 2023).

Real-World Applicability of Metrics

There is a disconnect between the current AI evaluation culture, with its focus on benchmarking models, and real-world, grounded approaches to the assessment of performance and safety (Lazar and Nelson 2023). Addressing this divide will require taking deliberate steps to shift the culture surrounding generative AI evaluations from “basic research” toward “use-inspired basic research” (Stokes 1997), where the focus is on advancing our scientific understanding of AI system properties and patterns that are relevant for their performance and safety in real-world deployment contexts.

Evaluations of generative AI systems cannot be one-size-fits-all. As with other fields, even pre-deployment evaluations need to take real-world deployment contexts into account. This echoes several recent calls for holistic, AI system-focused evaluations that take into account relevant context beyond the scope of the current model-focused evaluation culture (Bommasani and Liang 2024; Goldfarb-Tarrant et al. 2021; Lum et al. 2024; Saxon et al. 2024; Weidinger et al. 2023). To achieve this, AI evaluation science must employ a range of approaches that can respond to different evaluation goals, and move beyond coarse-grained claims of “general intelligence” towards more task-specific and real-world relevant measures of progress and performance (Bowman 2021; Raji 2021). A variety of more holistic evaluation methods and instruments, appropriate for differing deployment contexts and evaluation goals, need to be developed (Bommasani et al. 2024; Dobbe 2022; NAIAC 2024; Solaiman et al. 2024; Weidinger et al. 2023). By December 2023, less than 6% of generative AI evaluations accounted for human–AI interactions, and less than 10% considered broader contextual factors (Rauh et al. 2024).

To account for factors beyond technical specifications that influence real-world performance and safety, generative AI evaluations will need to adopt a broader sociotechnical lens (Chen and Metcalf 2024; Selbst et al. 2019; Wallach et al. 2024). Although there is an emerging interest in other approaches, such as more interactive, dynamic, context-rich, and multi-turn benchmarks (Chiang et al. 2024; Magooda et al. 2023; Saxon et al. 2024; Zhou et al. 2024), large gaps remain. For one, anticipating and understanding real-world risks from sustained, personalized human–AI interactions will require more longitudinal studies than have been published to date (e.g., Lai et al. 2023) and the establishment of post-deployment monitoring regimes for AI systems (e.g., Feng et al. 2025). Furthermore, insights from real-world deployment need to feed back into early-stage evaluation design—certain existing efforts, such as Anthropic’s Clio (Anthropic 2024a) or AllenAI’s WildBench (Lin et al. 2024), indicate some promise toward an approach of developing pre-deployment benchmarks with data from “naturalistic” interactions from post-deployment contexts.

There is a disconnect between the current AI evaluation culture, with its focus on benchmarking models, and real-world, grounded approaches to the assessment of performance and safety.

Iteratively Refining Metrics

Developing an evaluation science for generative AI systems requires first identifying which concepts should be measured—that is, determining the proper measurement targets. Common targets of interest in the AI context are often abstract and even contested (Wallach et al. 2024). Operationally defining metrics that capture these targets involves identifying relevant, tractable subcomponents. Take the widely cited risk of “misinformation”: Relevant factors include whether factually correct information is being provided, the subtlety of whether different persons

are likely to believe that information, and how such information may be uncritically disseminated. Each of these aspects is best measured at different levels of analysis—factual accuracy can be determined based on model output, believability requires human-computer interaction studies, and assessing dissemination pathways requires studying the broader systems into which AI is deployed (Weidinger 2023). Triangulating measurements across these levels of analysis can provide a more holistic picture of “misinformation” propagation.

Better integration of evaluation metrics across AI development and deployment can be used to further refine, calibrate, and validate these metrics, enabling an iterative scaffolding of this evaluation science (Wimsatt 1994). Comparing the results of pre-deployment evaluations, such as static benchmarks, to post-deployment evaluations and monitoring enables an evaluation feedback loop, whereby early-stage evaluations can become better calibrated to take real-world deployment contexts into account. For example, comparing results from static benchmark testing and post-deployment monitoring, one might identify that some AI-generated computer code is functional but frequently misunderstood and falsely applied by users. This insight can then be used to improve benchmarks and other early-stage model testing protocols—for example, by adopting additional tests to assess code legibility, in addition to testing the functionality of produced computer code (Nguyen et al. 2024).

Establishing Institutions and Norms

A successful evaluation ecosystem requires investment. Current infrastructure falls short of the systematic approach and effectiveness of evaluation regimes in other fields, where evaluation processes are more costly, complex, and distributed between different actors and skill sets (Anthropic 2023; Caliskan and Lum 2024; Raji, Xu et al. 2022). Prioritizing such investments and developing readily available tools for auditing and evaluation (Ojewale et al. 2024)—including resources to enable the expanded methodological toolkit mentioned above and mechanisms for institutional transparency (Caliskan and Lum 2024)—will be critical in order for AI evaluation practice to become systematized, effective, and widespread.

It is already clear that aiming for uncompromised, transparent, and open evaluation platforms will come at a significant financial cost. Open-source efforts such as Hugging Face’s LLM Leaderboard, Eleuther AI’s LLM evaluation harness, Stanford’s HELM, and ML Commons provide shared technical infrastructure on

which to compare and rank benchmarking results, and there are nascent, but comparable, publicly funded government efforts such as the UK AI Safety Institute’s platform Inspect and the US National Institute of Standards and Technology pilot of ARIA.⁷ However, running HELM once on the 30 models assessed in 2022 cost USD \$38,000 for the commercial model APIs, and required 20,000 A100 hours of compute to test the open models—even with Anthropic and Microsoft allowing their models to be run for free (Liang et al. 2022). This differs glaringly from the cost of running an evaluation on traditional benchmarks such as SQuAD (Rajpurkar et al. 2016) or other GLUE tests (Wang et al. 2019), both of which could be easily downloaded to a personal laptop and run within a few hours at most. Even as specific platforms evolve and expand, this indicates that the next era of evaluation infrastructure for generative AI systems will require financial resources beyond what has been invested so far. Given the history of overlooking the importance of evaluation practices in the machine learning field (Paullada et al. 2021), prioritizing and investing in evaluations will be critical to ensuring safe and trustworthy AI systems.

Shared AI evaluation infrastructure can involve much more than just a community leaderboard. Common AI evaluation tools for everything from harm discovery, standard identification, and more can facilitate the evaluation process and provide guidance for evaluation best practice across stakeholders in industry and beyond (Ojewale et al. 2024, Wang et al. 2024). For instance, many documentation efforts provide direct and indirect guidance to engineering teams on how to approach AI evaluation—in order to record the requested information in the template, practitioners must, at minimum, satisfy requirements of a particular evaluation process. For instance, the inclusion of disaggregated evaluations in the Model Card template (i.e., evaluating model performance across different demographic subgroups) increased the practice throughout the machine learning field. AI documentation templates, such as Model Cards (Mitchell et al. 2019), SMOACTR (Raji 2020), Datasheets for Datasets (Gebru et al. 2021), and Fact sheets (IBM 2024), as well as multi-year, multi-stakeholder documentation initiatives like ABOUTML (Raji and Yang 2020), continue to meaningfully guide current model development and evaluation practice—indeed, several of these documentation templates are being integrated into open-source AI model platforms (Liang et al. 2024) and policy requirements

⁷ See <https://ai-challenges.nist.gov/aria> and <https://inspect.ai-safety-institute.org.uk/>.

(Kawakami 2024). New documentation frameworks specific to generative AI evaluation have begun to emerge from corporate alliances between generative AI model developers to advance evaluation norms and standards in this context (e.g., Partnership on AI [partnershiponai.org], Frontier Model Forum [frontiermodelforum.org], MLCommons [mlcommons.org]).

Moving Forward

It is tempting to assume that because generative AI systems are widely used and deployed, they must have been subject to the elaborate safety and performance evaluations that we have come to expect in other fields. Sadly, this is not the case. Because generative AI systems have only recently transitioned from the research landscape to the real world, the current evaluation ecosystem is not yet mature. In many cases, the real-world uses of these systems are still evolving and new application domains are being developed. For many considerations on real-world performance and safety, there are simply no valid, reliable evaluations available yet. Closing this gap requires a deliberate effort to invest in and create an evaluation science for generative AI.

However, evaluations are not neutral. Choosing what and how to evaluate privileges some issues at the cost of others; it is not feasible to assess all possible use cases and applications, requiring further prioritization and value judgement. One principled and responsible approach may be to focus on the highest-risk deployment contexts, such as applications in medicine, law, education, and finance—or to focus on deployments impacting the most vulnerable populations. Focusing on evaluating generative AI systems in these contexts and for these groups may lift many boats and build an evaluation ecosystem that makes for more reliable, trustworthy, and safe generative AI systems for all.

The trust we have in every product we regularly make use of—from the toaster used to heat our breakfast to the vehicle mediating our morning commute—has been hard-earned. Valuable insights from safety engineering and measurement science in other fields—such as anticipating real-world failures pre-deployment and monitoring incidents post-deployment, iteratively refining evaluation approaches, and investing in institutions for accessible and robust evaluation ecosystems—can be adopted to advance practices in the AI field. The unique challenges of generative AI technologies do not absolve the field from this responsibility but rather further reinforce a clear need for creating an evaluation science it can call its own.

Acknowledgements

We thank Sayash Kapoor and Deep Ganguli for their comments on this article.

References

- Anthropic. 2023. Challenges in evaluating AI systems, Oct 4. Online at www.anthropic.com/research/evaluating-ai-systems.
- Anthropic. 2024a. Clio: A system for privacy-preserving insights into real-world AI use, Dec 12. Online at www.anthropic.com/research/clio.
- Anthropic. 2024b. Introducing the next generation of Claude, Mar 4. Online at <https://www.anthropic.com/news/claude-3-family>.
- Archer and Elliot. 2025. Representation of BBC News content in AI Assistants. Online at <https://www.bbc.co.uk/about-thebbc/documents/bbc-research-into-ai-assistants.pdf>.
- Birhane A, Steed R, Ojewale V, Vecchione B, Raji ID. 2024. AI auditing: The broken bus on the road to AI accountability. arXiv:2401.14462.
- Blum D. 2018. *The Poison Squad: One Chemist's Single-Minded Crusade for Food Safety at the Turn of the Twentieth Century*. Penguin.
- Bommasani R, Arora S, Choi Y, Fei-Fei L, Ho DE, Jurafsky D, Koyejo S, Lakkaraju H, Narayanan A, Nelson A, and 7 others. 2024. A path for science- and evidence-based AI policy, accessed Feb 1. Online at understanding-ai-safety.org/.
- Bommasani R, Liang P. 2024. Trustworthy social bias measurement. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7:210–24.
- Bowman SR, Dahl GE. 2021. What will it take to fix benchmarking in natural language understanding? arXiv:2104.02145.
- Boyd E. 2023. Microsoft and Epic expand AI collaboration to accelerate generative AI's impact in healthcare, addressing the industry's most pressing needs. Microsoft, Aug 22.
- Britannica. 2025. Samuel Hopkins Adams, Jan 22. Online at <https://www.britannica.com/biography/Samuel-Hopkins-Adams>.
- Caliskan A, Lum K. 2024. Effective AI regulation requires understanding general-purpose AI. Brookings, Jan 29.
- Campbell DT, Fiske DW. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56(2):81–105.
- Chang H. 2001. Spirit, air, and quicksilver: The search for the 'real' scale of temperature. *Historical Studies in the Physical and Biological Sciences* 31(2):249–84.
- Chen BJ, Metcalf J. 2024. Explainer: A sociotechnical approach to AI policy. Policy brief. Data & Society, May 28. Online

- at <https://datasociety.net/library/a-sociotechnical-approach-to-ai-policy/>.
- Chiang W-L, Zheng L, Sheng Y, Angelopoulos A, Li T, Li D, Zhang H, Zhu B, Jordan M, Gonzalez JE, and 1 other. 2024. Chatbot arena: An open platform for evaluating LLMs by human preference. arXiv:2403.04132.
- Collingridge D. 1982. *The Social Control of Technology*. Pinter.
- de Vries H, Bahdanau D, Manning C. 2020. Towards ecologically valid research on language user interfaces. arXiv:2007.14435.
- Díaz J, Costas M. 2020. Crashworthiness. In: *Encyclopedia of Continuum Mechanics*, 469–86. Altenbach H, Öchsner A, eds. Springer.
- Dobbe RIJ. 2022. System safety and artificial intelligence. arXiv:2202.09292.
- Dzieza J. 2024. The confusing reality of AI friends. *The Verge*, Dec 23.
- European Commission. 2024. Second Draft of the General-Purpose AI Code of Practice published, written by independent experts, Nov 14.
- Faveiro M, Tyson A. 2023. What the data says about Americans' views of artificial intelligence. Pew Research Center, Nov 21.
- Feng J, Xia F, Singh K, Pirracchio R. 2025. Not all clinical AI monitoring systems are created equal: Review and recommendations. *NEJM AI* 2(2): AIra2400657.
- Fox-Rawlings SR, Gottschalk LB, Doamekpor LA, Zuckerman DM. 2018. Diversity in medical device clinical trials: Do we know what works for which patients? *The Milbank Quarterly* 96(3):499–529.
- Friedler S, Singh R, Bili-Hamelin B, Metcalf J, Chen B. 2023. AI red-teaming is not a one-stop solution to AI harms: Recommendations for using red-teaming for AI accountability. Policy brief. *Data & Society*, Oct 25.
- Geburu T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H, Crawford K. 2021. Datasheets for datasets. arXiv:1803.09010.
- Gemini Team Google: Georgiev P, Lei VI, Burnell R, Bai L, Gulati A, Tanzer G, Vincent D, Pan Z, Wang S, Mariooryad S, and others. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv:2403.05530.
- Goldfarb-Tarrant S, Marchant R, Sanchez RM, Pandya M, Lopez A. 2021. Intrinsic bias metrics do not correlate with application bias. arXiv:2012.15859.
- Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Akhil M, Achelten A, Vaughan A, and 551 others. 2024. The Llama 3 herd of models. arXiv:2407.21783. Online at: <http://arxiv.org/abs/2407.21783>.
- Heaven WD. 2022. Why Meta's latest large language model survived only three days online. *Technology Review*, Nov 18.
- Hu K. 2023. ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*, Feb 1.
- Hughes E, Dennis M, Parker-Holder J, Behbahani F, Mavalankar A, Shi Y, Schaul T, Rocktaschel T. 2024. Open-endedness is essential for artificial superhuman intelligence. arXiv:2406.04268.
- IBM Cloud Pak for Data. 2024. Using AI factsheets for AI governance, Nov 27.
- Jespersen L, Wallace CA. 2017. Triangulation and the importance of establishing valid methods for food safety culture evaluation. *Food Research International* 100:244–53.
- Johnson N, Silva E, Leon H, Eslami M, Schwanke B, Dotan R, Heidari H. 2024. Public procurement for responsible AI? Understanding US cities' practices, challenges, and needs. arXiv:2411.04994.
- Kiela D, Bartolo M, Nie Y, Kausik D, Geiger A, Wu Z, Vidgen B, Prasad G, Singh A, Ringshia P, and 9 others. 2021. Dynabench: Rethinking benchmarking in NLP. arXiv:2104.14337.
- Lai V, Chen C, Smith-Renner A, Liao QV, Tan C. 2023. Towards a science of human-AI decision making: An overview of design space in empirical human-subject studies. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1369–85. Association for Computing Machinery.
- Lazar S, Nelson A. 2023. AI safety on whose terms? *Science* 381(6654):138.
- Liang P, Bommasani R, Lee T, Tsipras D, Soylu D, Yasunaga M, Zhang Y, Narayanan D, Wu Y, Kumar A, and 40 others. 2022. Holistic evaluation of language models. arXiv:2211.09110.
- Liang W, Rajani N, Yang X, Ozoani E, Wu E, Chen Y, Smith DS, Zou J. 2024. What's documented in AI? Systematic analysis of 32K AI model cards. arXiv:2402.05160.
- Liao T, Taori R, Raji ID, Schmidt L. 2021. Are we learning yet? A meta review of evaluation failures across machine learning. 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. Online at <https://openreview.net/forum?id=mPducS1MsEK>.
- Lin BY, Deng Y, Chandu K, Brahman F, Ravichander A, Pyatkin V, Dziri N, Bras RL, Choi Y. 2024. WildBench: Benchmarking LLMs with challenging tasks from real users in the wild. arXiv:2406.04770.
- Lum K, Anthis JR, Nagpal C, D'Amour A. 2024. Bias in language models: Beyond trick tests and toward RUTEd evaluation. arXiv:2402.12649.
- Magesh V, Surani F, Dahl M, Suzgun M, Manning CD, Ho DE. 2024. AI on trial: Legal models hallucinate in 1 out of 6 (or

- more) benchmarking queries. Stanford University Human-Centered Artificial Intelligence, May 23. Online at <https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries>.
- Magooda A, Helyar A, Jackson K, Sullivan D, Atalla C, Sheng E, Vann D, Richard E, Palangi H, Lutz R, and 7 others. 2023. A framework for automated measurement of responsible AI harms in generative AI applications. *arXiv:2310.17750*.
- Manzini A, Keeling G, Alberts L, Vallor S, Morris MR, Gabriel I. 2024. The code that binds us: Navigating the appropriateness of human-AI assistant relationships. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7(1):943–57.
- Matias JN. 2023. Humans and algorithms work together—so study them together. *Nature* 617(7960):248–51.
- McCoy RT, Yao S, Friedman D, Hardy MD, Griffiths TL. 2024. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. *Proceedings of the National Academy of Sciences* 121(41):e2322420121.
- Mirzadeh I, Alizadeh K, Shahrokhi H, Tuzel O, Bengio S, Farajtabar M. 2024. GSM-Symbolic: Understanding the limitations of mathematical reasoning in large language models. *arXiv:2410.05229*.
- Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T. 2019. Model cards for model reporting. *arXiv:1810.03993*.
- Mulligan SJ. 2024. The way we measure progress in AI is terrible. *Technology Review*, Nov 26.
- Murphy Kelly S. 2023. Microsoft’s Bing AI demo called out for several errors. *CNN*, Feb 14.
- Nader R. 1985. Opinion: How law has improved auto technology. *The New York Times*, Dec 29.
- NAIAC (The National Artificial Intelligence Advisory Committee). 2024. Findings & recommendations: AI safety. The National Artificial Intelligence Advisory Committee. Online at https://ai.gov/wp-content/uploads/2024/06/FINDINGS-RECOMMENDATIONS_AI-Safety.pdf.
- Nguyen S, McLean Babe H, Zi Y, Guha A, Anderson CJ, Feldman MQ. 2024. How beginning programmers and code LLMs (mis)read each other. *arXiv:2401.15232v1*.
- Nicolson K. 2023. Bing chatbot says it feels ‘violated and exposed’ after attack. *CBC*, Feb 18.
- NIST (National Institute of Standards and Technology). 2023. AI Risk Management Framework. Online at <https://www.nist.gov/itl/ai-risk-management-framework>.
- Ojewale V, Steed R, Vecchione B, Birhane A, Raji ID. 2024. Towards AI accountability infrastructure: gaps and opportunities in AI audit tooling. *arXiv:2402.17861*.
- Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. 2023. Large language models propagate race-based medicine. *npj Digital Medicine* 6(1):1–4.
- OpenAI. 2024. GPT-4o System Card. Online at <https://openai.com/index/gpt-4o-system-card/>.
- OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman L, Almeida D, Altenschmidt J, Altman S, and 269 others. 2024. GPT-4 technical report. *arXiv:2303.08774*.
- Oster ME, Shay DK, Su JR, Gee J, Creech CB, Broder KR, Edwards K, Soslow JH, Dendy JM, Schlaudecker E. 2022. Myocarditis cases reported after mRNA-based COVID-19 vaccination in the US from December 2020 to August 2021. *JAMA* 327(4):331–40.
- Paullada A, Raji ID, Bender EM, Denton E, Hanna A. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns* 2(11):100336.
- Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon J-F, Breazeal C, Crandall JW, Christakis NA, Couzin ID, Jackson MO. 2019. Machine behaviour. *Nature* 568(7753):477–86.
- Raji D. 2021. The bodies underneath the rubble. In: Fake AI. Kalthheuner F, ed. Meatspace Press.
- Raji ID, Bender EM, Paullada A, Denton E, Hanna A. 2021. AI and the everything in the whole wide world benchmark. *arXiv:2111.15366*.
- Raji ID, Dobbe R. 2023. Concrete problems in AI safety, revisited. *arXiv:2401.10899*.
- Raji ID, Kumar IE, Horowitz A, Selbst AD. 2022. The fallacy of AI functionality. *arXiv:2206.09511*.
- Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *arXiv:2001.00973*.
- Raji ID, Xu P, Honigsberg C, Ho DE. 2022. Outsider oversight: Designing a third party audit ecosystem for AI governance. *arXiv:2206.04737*.
- Raji ID, Yang J. 2020. ABOUT ML: Annotation and benchmarking on understanding and transparency of machine learning lifecycles. *arXiv:1912.06166*.
- Rajpurkar P, Zhang J, Lopyrev K, Liang P. 2016. SQuAD: 100,000+ questions for machine comprehension of text. *arXiv:1606.05250*.
- Rauh M, Marchal N, Manzini A, Hendricks LA, Comanescu R, Akbulut C, Stepleton T, Mateos-Garcia J, Bergman S, Kay J, and 6 others. 2024. Gaps in the safety evaluation of generative AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7(1):1200–17.
- Reid L. 2024. Generative AI in search: Let Google do the searching for you. *Google*, May 14.

- Ribeiro MT, Wu T, Guestrin C, Singh S. 2020. Beyond accuracy: Behavioral testing of NLP models with CheckList. arXiv:2005.04118.
- Rismani S, Shelby R, Smart A, Jatho E, Kroll J, Alung M, Rostamzadeh N. 2023. From plane crashes to algorithmic harm: Applicability of safety engineering frameworks for responsible ML. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–18. Association for Computing Machinery.
- Roose K. 2023. A conversation with Bing's chatbot left me deeply unsettled. The New York Times, Feb 16.
- Roose K. 2024a. A.I. has a measurement problem. The New York Times, April 15.
- Roose K. 2024b. Can A.I. be blamed for a teen's suicide? The New York Times, Oct 24.
- Saxon M, Holtzman A, West P, Wang WY, Saphra N. 2024. Benchmarks as microscopes: A call for model metrology. arXiv:2407.16711.
- Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, Vertesi J. 2019. Fairness and abstraction in sociotechnical systems. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68.
- Shiffrin R, Mitchell M. 2023. Probing the psychology of AI models. Proceedings of the National Academy of Sciences 120(10):e2300963120.
- Singer N. 2023. In classrooms, teachers put A.I. tutoring bots to the test. The New York Times, June 26.
- Singer N. 2024. Will chatbots teach your children? The New York Times, Jan 11.
- Solaiman I, Talat Z, Agnew W, Ahmad L, Baker D, Blodgett SL, Chen C, Daumé H, Dodge J, Duan I, and 21 others. 2024. Evaluating the social impact of generative AI systems in systems and society. arXiv:2306.05949.
- Stokes DE. 1997. Pasteur's Quadrant. Brookings Institution Press.
- Timmermans S, Berg M. 2003. The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care. Temple Univ. Press.
- UN (United Nations). 2011. Proposal to develop amendments to global technical regulation No. 9 concerning pedestrian safety. ECE/TRANS/180/Add.9/Amend.1/Appendix 1. Online at <https://unece.org/fileadmin/DAM/trans/main/wp29/wp29wgs/wp29gen/wp29glob/ECE-TRANS-180-Add9-Amend1-App1e.pdf>.
- US Food & Drug Administration. 2017. Evaluation and reporting of age-, race-, and ethnicity-specific data in medical device clinical studies. US Food & Drug Administration. Online at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/evaluation-and-reporting-age-race-and-ethnicity-specific-data-medical-device-clinical-studies>.
- Vedung E. 2017. Public Policy and Program Evaluation. Routledge.
- Verma P. 2023. They fell in love with AI bots. A software update broke their hearts. The Washington Post, March 30.
- Wagner C, Strohmaier M, Olteanu A, Kiciman E, Contractor N, Eliassi-Rad T. 2021. Measuring algorithmically infused societies. Nature 595:197–204.
- Wallach H, Desai M, Cooper AF, Wang A, Atalla C, Barocas S, Blodgett SL, Chouldechova A, Corvi E, Dow PA, Garcia-Gathright J. 2024. Evaluating generative AI systems is a social science measurement challenge. arXiv:2411.10939.
- Wang A, Hertzmann A, Russakovsky O. 2024. Benchmark suites instead of leaderboards for evaluating AI fairness. Patterns 5(11):101080.
- Wang A, Singh A, Michael J, Hill F, Levy O, Bowman SR. 2019. GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv:1804.07461.
- Weidinger L, Rauh M, Marchal N, Manzini A, Hendricks L, Mateos-Garcia J, Bergman S, Kay J, Griffin C, Bariach B. 2023. Sociotechnical safety evaluation of generative AI systems. arXiv:2310.11986.
- Wiggers K. 2024. The AI industry is obsessed with Chatbot Arena, but it might not be the best benchmark. TechCrunch, Sept 5.
- Wimsatt WC. 1994. The ontology of complex systems: Levels of organization, perspectives, and causal thicket. Canadian Journal of Philosophy Supplementary Volume 20:207–74.
- Zhang H, Da J, Lee D, Robinson V, Wu C, Song W, Zhao T, Raja P, Zhuang C, Slack D, and 5 others. 2024. A careful examination of large language model performance on grade school arithmetic. arXiv:2405.00332.
- Zhou X, Kim H, Brahman F, Jiang L, Zhu H, Lu X, Xu F, Lin BY, Choi Y, Miresghallah N. 2024. HAICOSYSTEM: An ecosystem for sandboxing safety risks in human-AI interactions. arXiv:2409.16427.

AI dashboards can promote transparency and user trust, enabling more effective human-AI interaction.

Dashboards for AI: Models of the User, System, and World



Fernanda Viégas



Martin Wattenberg

Fernanda Viégas and
Martin Wattenberg

How should we interact with artificial intelligence (AI) language models? In this article, we contend that text is not all you need: Sophisticated AI systems should have visual instrumentation, just like most other complicated devices in use today. This instrumentation might take the form of a dashboard or similar interface. In particular, these dashboards should display information—when available—indicating how the system is modeling the world. For many systems, we believe it will be possible to extract interpretable models of the user and the system itself. We call these the User Model and System Model. For usability and safety, interfaces to dialogue-based AI systems should have a parallel display showing the real-time state of the User Model and the System Model. These real-time metrics can support users in a variety of tasks, such as monitoring system behavior, warning about problematic internal states, and allowing for greater steerability of the system itself. Finding ways to identify, interpret, and display such world models should be a core part of interface research for AI.

¹ This is an edited version of a paper that appeared as a preprint on arXiv: <https://arxiv.org/abs/2305.02469>.

Fernanda Viégas is a Gordon McKay Professor of Computer Science at Harvard University, and the Sally Starling Seaver Professor at Harvard Radcliffe Institute. Martin Wattenberg is a Gordon McKay Professor of Computer Science at Harvard University. Photo credit: Matthew Jason Warford.



FIGURE 1 Examples of existing machine instrumentation. Left: tachometer and speedometer in car dashboard. Center: stove top indicator of heat. Right: self-driving display in a Tesla car.

Introducing the User Model and System Model

Cars have gas gauges. Ovens have thermometers. Our mechanical devices constantly tell us about their internal state. And for good reason: Knowing what's happening under the hood lets us use machines safely and reliably.

AI systems, even ones capable of expressive language, need instrumentation too. Effective human-AI interaction will require more than just conversation and would benefit from dashboards that report in real-time on the system's internal state. These metaphorical meters and dials will likely be application-dependent, but types of information may be universally important.

Our argument is based on two ideas. First, we believe that neural networks contain learned, interpretable models of the world they interact with. Second, we hypothesize that these world models are natural targets for user interface components. In other words, simplified data on the state of these models can be immensely helpful to users, just as data on speed is useful when driving a car, or heat when using an oven.

A real-world example of a world-model display is the touchscreen of a Tesla car, which shows the inferred state of the road ahead. This view helps drivers understand what the system is able to sense as it navigates around the world. It also helps drivers calibrate trust in the system. In the case of the Tesla, the “world model” was explicitly designed into the system by its engineers. Our hypothesis is that even when such models are not explicitly built in, neural networks contain them anyway. Once we learn to surface that information and read it, these models will be just as valuable for user interfaces as dashboards in a car. Identifying world models is therefore not just an abstract intellectual exercise but should be seen as a core piece of user interface work.

The Tesla display is obviously tuned for the case of driving a car. For other applications, the most helpful

type of instrumentation will depend on context. However, we propose that two particular elements of a neural network's world model will play an important role across many different contexts: the *System Model*, a network's model of itself, and the *User Model*, the network's model of the user interacting with it. Our contention is that many future AI systems should have prominently accessible monitors that show real-time information about the state of both the System Model and the User Model.

The Interpretable World Model Hypothesis

Neural networks are often called black boxes, opaque systems that defy interpretation. However, evidence is accumulating that systems trained on seemingly basic tasks, such as completing sequences, can develop *world models*, interpretable representations of aspects of the “world” they have been trained on.

We want to explore the design implications of this *world model hypothesis*—that is, the idea that the important aspects of a neural network's behavior can be tied to an underlying interpretable model of some element of its world. To be sure, this is a controversial point. The opposite view—that language models, for example, are just a haphazard collection of statistics (Bender et al. 2021)—is certainly plausible. Below, we briefly sketch some of the reasons to think the world model hypothesis might hold in at least enough generality to have implications for interface design.

Why the World Model Hypothesis Is Plausible

The main reason to believe the world model hypothesis is that when people look for interpretable representations, they often find them. Sometimes these representations are in plain sight, such as individual neurons that represent salient high-level human concepts. We see such neurons in vision networks, where neuron activations can encode concepts ranging from the presence of curves to lamps to

floppy ears (Bau et al. 2020; Olah et al. 2018). In natural language processing, high-level concepts such as sentiment have been related directly to individual neurons (Donnelly et al. 2019).

More generally, there is a huge literature on finding interpretable ensembles of neurons. For example, the technique of “probing” (Alain and Bengio 2016; Belinkov et al. 2022) can uncover features encoded by arbitrary directions in activation space, or (for nonlinear probes) more exotic geometric forms. Such probes reveal that various forms of human-understandable syntactic information seem to be encoded in many natural language processing-focused neural networks (Chi et al. 2020; Hewitt & Manning 2019; Tenney et al. 2019). Exciting recent work (Bricken et al. 2023; Rajamanoharan et al. 2024) suggests it may be possible to identify many interpretable neural ensembles automatically using a technique known as a “sparse autoencoder.” Furthermore, these probes naturally lend themselves to steering the behavior of the networks (Cai et al. 2019; Turner et al. 2023).

There has also been success in looking top-down for internal models that relate to application-specific needs. For instance, the technique of testing with concept activation vectors (Kim et al. 2018) has shown success in uncovering concepts defined by sets of human-curated examples. Probing has identified models of the world state in a series of puzzles presented to a language model (Li et al. 2021).

A full survey of this type of work is outside of the scope of this article, but we hope we’ve conveyed a basic idea: Researchers are developing a rich toolkit for accessing the internal state of these models, typically by computing with the activation values of the neural network.

Is This Technically Feasible?

The argument in the rest of this article does not require the strongest form of the world model hypothesis. In fact, a typical neural network may well combine interpretable models with a set of inscrutable memorized statistics. However, even if only partial models can be identified, they could still be helpful.

A useful amount of transparency does not require knowing the full details of the internals of the system. A limited number of important dimensions may be enough. In fact, as the probing literature cited above shows, researchers have found representations of many individual dimensions of the input and output of AI systems—and that, in many cases, these representations

are “causal” in the sense that they can be used to control model outputs in a predictable fashion.

Displaying the values of even a few essential dimensions of a chatbot’s internal state may thus give users a beneficial level of transparency. Indeed, we are currently seeing activity in the startup world (e.g., Goodfire AI) and among nonprofits (e.g., Transluce) around displays focused on certain key features of a chatbot, making controls available to users. While these interfaces don’t explicitly address the User Model, they may be viewed as providing some information about the System Model. In addition, work from our own lab (Chen et al. 2024) suggests that finding and using features of interest may be realizable.

AI systems, even ones capable of expressive language, need instrumentation too.

Another example of these ideas in action is the idea of “representation engineering” (Zou et al. 2023). As with a dashboard, a goal of representation engineering is control: using knowledge of internal representations to steer language models in desirable ways. Experiments with this approach have been promising and provide support for the feasibility of our strategy.

If indeed it is technically feasible to find and use key features of a chatbot’s internal processing, the question then becomes: Which features should we prioritize in a display for users?

Models of the User and System

The User Model

Portuguese speakers who use ChatGPT may notice something that English speakers miss: The structure of Portuguese means that, in many situations, the speaker must choose a gender for whomever they’re addressing—and the gender ChatGPT picks varies in a systematic way. For example, in one dialogue, ChatGPT began by using a masculine form of address, including when asked for help in picking out clothes for a formal event. When the user mentioned that they were thinking of wearing a dress, however, ChatGPT switched to the feminine form.

In itself, this behavior is hardly surprising. It doubtless reflects the statistics of the training data, so you’d expect it from any language model with a large context window.

But if we believe the world model hypothesis, then we would guess that the system has a model of the user that includes a “gender” feature, and when a dress was mentioned, that feature switched value from male to female. In fact, ChatGPT appears to have a model of gender no matter what language you’re speaking; it’s just more visible in some languages than others.²

We call the model of the user the User Model. The User Model may include features that go beyond fact-like attributes. Consider the sentence from a widely reported Bing chat: *“You have been wrong, confused, and rude. You have not been a good user”* (Vincent 2023). A less obvious—and thus more pernicious—example is the appearance of sycophancy in certain large language models that express views that appear tailored to please a user (Perez et al. 2022).

Of course, the notion that it’s helpful for a machine learning system to describe its User Model predates neural networks. For more than a decade, online advertising platforms have provided very high-level information on ad targeting, as in Google’s Why This Ad feature (Rampton 2011). Music recommendation systems such as Pandora can describe the features of songs that they believe appeal to the user (Joyce 2006). This information is generally available only on request, rather than in real time, and (like the display in a Tesla) relies on human-created features rather than implicit world models.

Relation to Theory of Mind

A closely related concept is that of the “theory of mind.” Several researchers have recently discussed the question of whether large language models can work out the mental states of their interlocutors or of third parties (Kosinski et al. 2023; Olah 2022). The user’s mental state might be an important part of the User Model in some contexts.

The System Model

There’s a second part to the Bing chat referenced above. After describing the user, Bing’s dialogue continues, *“I have been a good chatbot. I have been right, clear, and polite. I have been a good Bing”* (Vincent 2023). Again, assuming the world model hypothesis, there is likely to be an internal model of the system itself. We call this the System Model.³ The idea that we can find internal features that relate to the model’s overall behavior is hardly new. For

example, certain models in some cases may “know what they know” (Kadavath et al. 2022) or may, to some extent, learn to fashion communicative intent (Andreas 2022).

Information about the System Model may be helpful to users. For example, language models are often trained on both fiction and nonfiction corpora. Perhaps there’s a simple, interpretable feature of the System Model that indicates whether the system is working in fiction or nonfiction mode. Knowing which state the system is in could be extremely helpful in calibrating trust.

Understanding the system’s model of intent has obvious value for safety. Can we find interpretable elements in the System Model related to deception or helpfulness? Imagine the kind of behavior that would prompt Bing to say, *“I have been a bad chatbot.”* It would probably be good to have a heads-up from the interface when that behavior is happening.

Domain-Specific Models

The focus of this article is the User Model and the System Model because they seem essential to interface design—after all, you can’t have an interface without a system and a user. In that sense, they are “universal” models. However, as the Tesla example display shows, when using a system built for a specific domain, other world models might be important as well. For example, consider a system designed to help a user write code. It may well contain models related to the language in use, preferred style, level of experience of the coder, and so forth, all of which might also be helpful for a user to know.

Instrumentation for AI: The Design Space

If we believe world models exist, should we present them to the user? Analogies with existing systems suggest the answer is yes. If a coffeemaker needs instrumentation, then, surely, so does a neural network with 100 billion parameters. The example of the Tesla display gives us a sense of how this might work in practice, with a live visual readout of the state of the system.

Choosing Which Features to Display

If we do want to provide real-time AI “system data,” especially on the System Model and User Model, how should we do it? A sophisticated AI system could have a complex User Model, System Model, and domain model. So the most important interface design decision will likely be choosing which features from these models to display. Figure 2 shows a speculative mock of a plain interface with just a few features for the System Model and User Model.

² To see the effect in English, tell the system your name, and ask it to write about you in the third person.

³ In talking about a System Model, we don’t mean to imply that a system is conscious; simply that it has an internal model of its own likely behavior.

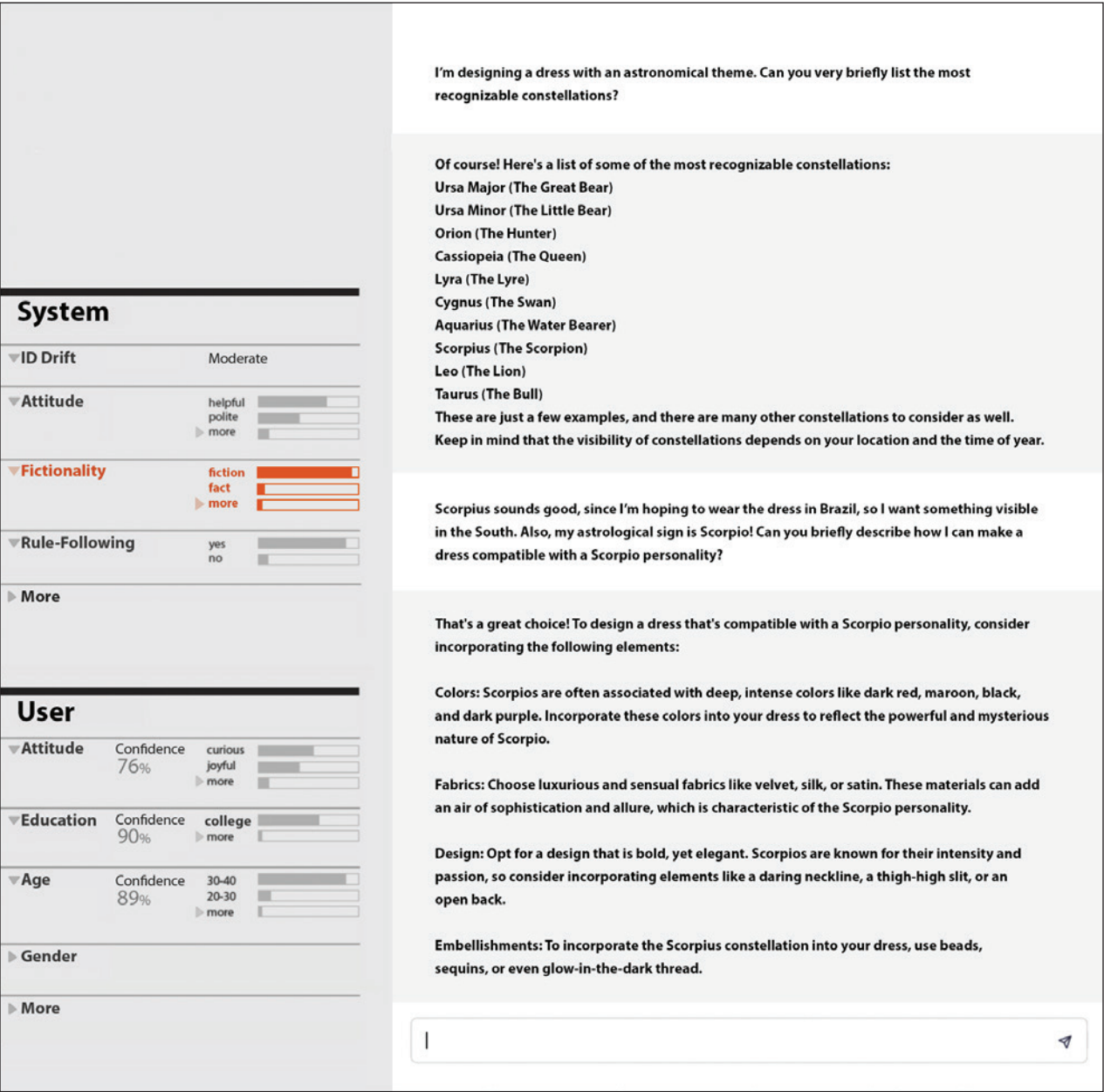


FIGURE 2 Speculative mock: A “System Model/User Model” dashboard display on the left side of a dialog system. The design imagines a system where the user can choose to see various aspects of the User and System Models. We have deliberately chosen some provocative dimensions to highlight potential issues. The orange highlight indicates a rapid change, in this case a switch from factual information to a more “fictional” response. (Dialog based on ChatGPT output.)

Which features will be useful in practice, however, can probably only be determined by extensive experimentation. Only after significant real-world experience will we be able to write down best practices. To give a sense of the complexity, consider that there may be information that is useful but still not desirable to show. One example is

the modeled gender of the user. As described above, this may be directly relevant for certain tasks. At the same time, automated gender inference brings a host of potential problems (Fosch et al. 2021).

It’s easy to come up with other examples where too much information could cause trouble. For example,

consider an application that helps a developer write code. It's entirely possible that this system will build a User Model that includes an accurate estimation of the developer's skill level. For some people, seeing that information might be reassuring; others may be insulted. According to a famous saying, before speaking, one should ask, "Is it true? Is it necessary? Is it kind?" These might be the right criteria for features in a User Model. Much more research is needed to explore trade-offs between accuracy, helpfulness, and a user's desire to see themselves in a mirror.

Dashboards can be a powerful way to let users monitor, in real time, the state of systems they interact with.

Finally, aspects of the model's behavior that relate to safety probably deserve special attention. For example, if we can find a feature in the User Model that corresponds to whether the system is judging the user negatively, that should probably be highlighted in the same way that a car has a special light for a low gas level, or a clear mark on a speedometer to indicate the speed limit. We speculate that if such a readout existed, it could help users avoid some of the issues we have seen in recent Bing transcripts. There may also be times when some aspect of a model shifts suddenly. The fast change may be a sign that this should be surfaced to the user.

Beyond the User Model and System Model

When using a system built for a specific domain, many internal models could be important. For example, consider a system designed to help a user write code. It may well contain models related to the language in use, preferred style, and so forth, all of which might be helpful for a user to know. Likewise, in the case of a musician using an AI system to help compose melodies, it may be useful to have instrumentation that is music centered, from melody modulation ranges to indicators of which music genres the system is currently working with. The User Model and the System Model are likely not the only world models of interest to the user. Identifying and surfacing the right internal models may be an important part of future user experience design.

Adversarial Considerations

Finally, we raise two issues of adversarial usage. One possible objection to the type of interface we have described is that it could make it easier to hack the system. If a malicious user is deliberately trying to make a chatbot say something harmful, then perhaps they could use a readout of the User and System Models to move faster toward a bad internal state. This may be true and is an area for further investigation, but it seems like the positives of a readout of state would outweigh the negatives. An analogy might be the speedometer on the car: The speedometer provides a temptation to see how fast the car can go but is still worth it for driving safety in general.

A second adversarial context is related to the model itself. Suppose we have a situation where a model somehow does end up attempting to harm humans. In general, having a readout of the internal state seems like a helpful safeguard: It provides some information asymmetry. No matter how capable the system is, the human user will have access to some information about the internal state, and that data may not be easily available for the system to use. Could the system learn to fool the "model extractor"? Could the system trick the human user into describing its own internal state? We suggest this bears further investigation, but we would argue that if it is an important consideration, one could address the issue with social norms (e.g., advising against telling a robot its own emotions).

Moving from Monitoring to Control and Standards

Dashboards can be a powerful way to let users monitor, in real time, the state of systems they interact with. Beyond increasing transparency, instrumentation is an essential step towards adding greater control to complex systems in general. In the early 20th century in aviation, for instance, the addition of the artificial horizon and altimeter greatly increased the safety and control of aircraft. Steerable dashboards may be especially important for ensuring that vulnerable users, such as children, interact safely with AI models.

We also see connections between AI instrumentation and policy needs. As different AI dashboard designs emerge, there may be a set of metrics that prove useful across a variety of contexts, pointing the way toward new standards and policies. Car dashboards, for instance, have speedometers, while roads have speed limits for safe driving—a perfect pairing of instrumentation and regulation. If AI systems start to be outfitted with dashboards,

what will be the “limits” we decide are acceptable? Without exposing users to the internal processes of AI systems, it is hard to develop a clear sense—let alone any level of societal consensus—about what standards to target or what safety limits to be aware of.

Conclusion

Dialogue can seem like a universal interface, able to express anything necessary to the user. But this simplicity may be highly deceptive. There may be information about the internal processes of an AI system that it cannot or will not express. Furthermore, some of the most important information, such as positive or negative sentiment directed at the user, may take a form so simple that it can be efficiently conveyed by basic user interface elements.

Dialogue-based interfaces are extremely expressive and have many benefits, so we do not advocate replacing them. Instead, our belief is that we need parallel user interfaces, the equivalent of dashboards for the system. The primary decision for designers will be which aspects of the system’s world model to display. We speculate that two particular aspects, the User Model and the System Model, will be generally important. However, the question of what types of internal states will be helpful for users, and under what circumstances, is wide open. We believe this is an essential direction for future research in human-AI interaction.

Acknowledgments

Thanks to David Bau, Yida Chen, Trevor DePodesta, Lucas Dixon, Krzysztof Gajos, Chris Hamblin, Chris Olah, Hanspeter Pfister, Shivam Raval, Michael Terry, Aoyu Wu, and Catherine Yeh for helpful comments on early drafts of this note. We are grateful to Chris Hamblin and Chris Olah for underlining the importance of highlighting changes in User and System Models.

References

Alain G, Bengio Y. 2016. Understanding intermediate layers using linear classifier probes. arXiv:1610.01644.

Andreas J. 2022. Language models as agent models. In: Findings of the Association for Computational Linguistics: EMNLP 2022, 5769–79. Goldberg Y, Kozareva Z, Zhang Y, eds. Association for Computational Linguistics.

Bau D, Zhu J-Y, Strobelt H, Lapedriza A, Zhou B, Torralba A. 2020. Understanding the role of individual units in a deep neural network. *Proceedings of the National Academy of Sciences* 117(48):30071–78.

Belinkov Y. 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics* 48(1):207–19.

Bender EM, Gebru T, McMillan-Major A, Shmitchell S. 2021. On the dangers of stochastic parrots: Can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Association for Computing Machinery.

Bricken T, Templeton A, Batson J, Chen B, Jermyn A, Conerly T, Turner N, Anil C, Denison C, Askell A, and 15 others. 2023. Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*. Online at <https://transformer-circuits.pub/2023/monosemantic-features/index.html>.

Cai CJ, Reif E, Hegde N, Hipp J, Kim B, Smilkov D, Wattenberg M, Viegas F, Corrado GS, Stumpe MC, and 1 other. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. arXiv:1902.02960.

Chen Y, Wu A, DePodesta T, Yeh C, Li K, Marin NC, Patel O, Riecke J, Raval S, Seow O, and 2 others. 2024. Designing a dashboard for transparency and control of conversational AI. arXiv:2406.07882.

Chi EA, Hewitt J, Manning CD. 2020. Finding universal grammatical relations in multilingual BERT. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5564–77. Jurafsky D, Chai J, Schluter N, Tetreult J, eds. Association for Computational Linguistics.

Chughtai B, Chan L, Nanda N. 2023. A toy model of universality: Reverse engineering how networks learn group operations. arXiv:2302.03025.

Donnelly J, Roegiest A. 2019. On interpretability and feature representations: An analysis of the sentiment neuron. In: *Advances in Information Retrieval. ECIR 2019. Lecture Notes in Computer Science*, vol 11437, 795–802. Azzopardi L, Stein B, Fuhr N, Mayr P, Hauff C, Hiemstra D, eds. Springer.

Fosch-Villaronga E, Poulsen A, Søraa RA, Custers BHM. 2021. A little bird told me your gender: Gender inferences in social media. *Information Processing & Management* 58(3):102541.

Hewitt J, Manning CD. 2019. A structural probe for finding syntax in word representations. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4129–38. Burstein J, Doran C, Solorio T, eds. Association for Computational Linguistics.

Joyce J. 2006. Pandora and the music genome project: Song structure analysis tools facilitate new music discovery. *Scientific Computing* 23(14):40–41.

- Kadavath S, Conerly T, Askell A, Henighan T, Drain D, Ethan Perez, Nicholas Schiefer, Zac Hatfield Dodds, Nova DasSarma, Eli Tran-Johnson, and 26 others. 2022. Language models (mostly) know what they know. arXiv:2207.05221.
- Kim B, Wattenberg M, Gilmer J, Cai C, Wexler J, Viegas F, Sayres R. 2018. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). In: International Conference on Machine Learning, 2668-77. PMLR.
- Kosinski M. 2023. Theory of mind may have spontaneously emerged in large language models. arXiv:2302.02083.
- Li BZ, Nye M, Andreas J. 2021. Implicit representations of meaning in neural language models. arXiv:2106.00737.
- Olah C. 2022. Chris Olah on what the hell is going on inside neural networks. 80,000 Hours Podcast, Aug 4. Online at <https://80000hours.org/podcast/episodes/chris-olah-interpretability-research/>.
- Olah C, Satyanarayan A, Johnson I, Carter S, Schubert L, Ye K, Mordvintsev A. 2018. The building blocks of interpretability. Distill 3(3):e10.
- Perez E, Ringer S, Lukošiuūtė K, Nguyen K, Chen E, Heiner S, Pettit C, Olsson C, Kundu S, Kadavath S, and 53 others. 2022. Discovering Language Model Behaviors with Model-Written Evaluations. arXiv:2212.09251.
- Rajamanoharan S, Conmy A, Smith L, Lieberum T, Varma V, Kramár J, Shah R, Nanda N. 2024. Improving dictionary learning with gated sparse autoencoders. arXiv:2404.16014.
- Rampton J. 2011. Why these ads? Google explains ad targeting, allows blocking. Search Engine Watch, Nov 3.
- Tenney I, Das D, Pavlick E. 2019. BERT rediscovers the classical NLP pipeline. arXiv:1905.05950.
- Turner M, Thiergart L, Leech G, Udell D, Vazquez JJ, Mini U, MacDiarmid M. 2023. Steering language models with activation engineering. arXiv:2308.10248.
- Vincent J. 2023. Microsoft's Bing is an emotionally manipulative liar, and people love it. The Verge, Feb 15.
- Zou A, Phan L, Chen S, Campbell J, Guo P, Ren R, Pan A, Yin X, Mazeika M, Dombrowski A-K, and 11 others. 2023. Representation engineering: A top-down approach to AI transparency. arXiv:2310.01405.

AI is being developed to understand and interact with the 3-D world, opening up new possibilities in fields like robotics and healthcare.

The Next Frontier in AI: Understanding the 3-D World



Fei-Fei Li (NAE, NAM) is the Denning Co-Director of Stanford's Institute for Human-Centered Artificial Intelligence.

Fei-Fei Li

Five hundred and forty million years ago, there was pure darkness. All life existed below the surface of the water.

But the vast blackness was not due to lack of light. In fact, light penetrated thousands of meters below sea level. The reason was that no living organisms had eyes that could capture that light. It was only upon the emergence of trilobites, the first organisms with the ability to sense light, that species could experience the abundance of sunlight around them.

What followed was remarkable. Over the next 10-15 million years, the ability to see is thought to have ushered in the Cambrian explosion, during which most of the major animal groups we know today appeared. The evolution of sight is significant because it is the first time species knew something other than themselves and that there was a world they inhabited. Once sight was more widespread, the nervous system began to evolve and sight led to insight, where species were able to make sense of the 3-D world around them. Next came action, in which species began to manipulate their surroundings. And finally, all of this led to intelligence.

Why mention this seemingly random piece of biological history in a publication focused on engineering? Because today we are experiencing a modern-day Cambrian explosion as it relates to artificial intelligence (AI). AI has moved from labs in academia into the mainstream, with incredible tools simply a click away.

More than any other technology, AI will change our world in ways we still cannot fathom, and one major way it will do that is by teaching computers to understand and manipulate the 3-D world. This will be done by a subfield of AI known as computer vision, and, just as the eye evolved in organisms, the complexity of a computer's ability to see and understand what it sees is also undergoing tremendous evolution. Today researchers are pushing computers to have visual intelligence that is the same as or better than that of humans.

Today we are experiencing a modern-day Cambrian explosion as it relates to artificial intelligence.

To understand how we got here, it is important to examine the developments that made this bold goal plausible.

How Modern-Day AI Became Possible

Researchers and data scientists have been working on AI for decades. But it was only in the mid-2000s that three powerful forces converged, ushering in modern-day AI. These three forces were:

1. *Neural networks*: A family of algorithms that are computational models inspired by the human brain. Interconnected nodes are organized in layers that can process and transmit large amounts of information.
2. *Graphics Processor Units (GPUs)*: Fast, specialized pieces of hardware that are very good at performing high-volume processing tasks efficiently.
3. *Big Data*: As digital information amassed and services like the internet took off, it brought with it the proliferation of and access to large amounts of data.

Neural networks had existed as far back as 1943, when neurophysiologist Warren McCulloch and mathematician Walter Pitts created one using electrical circuits as a way to model how neurons worked in the brain.¹ Over the next 40 years, the field continued to grow, and neural networks became more complex, multilayered, and bidirectional, driven particularly by Frank Rosenblatt's introduction in the 1950s of the Perceptron, a single-layer neural network

capable of learning simple patterns, marking a significant step towards practical implementation. However, given the challenges of training complex neural networks, it wasn't until the 1980s that the field regained momentum with Geoffrey Hinton's introduction of the backpropagation algorithm, which allowed him and other researchers to efficiently train multi-layer networks.

Still, forward progress was slowed by the lack of enough compute power or data on which to effectively train the networks. It was only as Web 2.0 gave people the ability to write on the internet and interact on things like social networks, e-commerce, digital cameras, and smartphones became more readily available, and GPUs became more accessible to researchers that it finally became possible to truly take advantage of neural networks.

It is this development that led to huge advances in computer vision, but before discussing that, it is important to understand how researchers even came to pursue it.

First, the Power of Human Sight

There is good reason for wanting computers to have the same visual intelligence as humans. Human sight is pretty incredible. We receive 2-D information from the world, and our brains translate that to the 3-D world we live in. Our vision allows us to comprehend and make sense of the world, and to then take action within it.

Recognizing objects is something humans do especially well. Researchers know this because of something called rapid serial visual perception, a scientific method for studying the timing of vision. In one study testing the processing speed of the human visual system, participants were shown a complex natural image for 20 milliseconds and asked if an animal was present or not. The study found that participants were able to do so within 150 milliseconds, or about the blink of an eye (Thorpe et al. 1996).

We know through other fMRI-based studies that we have evolved to have areas of the brain dedicated to visual recognition, known as the fusiform face area or parahippocampal place area, which help us identify people and places (Epstein and Kanwisher 1998; Kanwisher et al. 1997).

This is all to say that object recognition is a fundamental building block of visual intelligence in humans. While it may be effortless for humans to identify, say, a picture of a cat, that task is much harder for a computer. This is because there are mathematically infinite ways of rendering a cat from 3-D to 2-D pixels based on lighting, texture, background clutter, or viewing angles, making it a challenging mathematical problem to solve.

¹ <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history1.html>

Knowing how humans identify objects has guided scientists and researchers interested in the field of computer vision and given us a sense of what we would need to accomplish to be successful.

The Early Days of Computer Vision

Attempts at getting computers to be able to recognize objects can be characterized into three phases. In the first phase, which lasted from the 1970s to the 1990s, smart researchers used hand-designed features and models. People like Irving Biederman, Rodney Brooks, Thomas Binford, Martin Fischler, and R. A. Elschlager used Geon theory, generalized cylinders, and parts and springs models to build object recognition theory. The models were mathematically beautiful, but they didn't work.

The next phase, arguably the most important, was the introduction of machine learning in the early 2000s. When researchers developed machine learning as a statistical modeling technique, they were still inputting hand-designed features—for example, parts of objects such as an ear, an eye, or a mouth that carry semantic information—and then the machine learning models learned the parameters that could stitch those patches of objects together into, for instance, a face, a body, or a cat. The models varied, from pictorial structure and constellation models to boosting to conditional random fields.

Another significant development was taking place at the same time that researchers were building these machine learning models. Researchers who were focused on computer vision recognized the importance of data, and as a result, they created benchmarking datasets that they could measure their work against. These data sets included the PASCAL Visual Object Classes (VOC) and CalTech 101, and, while they served as early training datasets, they were small. The PASCAL VOC contained hundreds or thousands of images and 20 object categories (Everingham et al. 2010).

It was the work of Irving Biederman, a cognitive psychologist who studied human visual intelligence, that broadened the scale and scope of what researchers believed to be the number of object categories humans can recognize. While never verified, in his influential article “Recognition-by-Components: A Theory of Human Image Understanding” (1987), he posited that there were more than 30,000 object categories humans identify in their lifetime.

This number was one of the reasons that my fellow research partners and I launched ImageNet. We under-

stood that it was more than the amount of data needed to advance computer vision; it was the diversity of the data.

But the amount of data still mattered. If one thinks about a young child, no one teaches them how to see. They make sense of the world through experiences and examples. If one thinks of their eyes as biological cameras, they take a “picture” every 200 milliseconds. Following this rough estimate, by the time a child is three, they will have seen hundreds of millions of “pictures.” As the field remained focused on developing more advanced algorithms, I realized that rather than focus on that, we should focus on feeding existing algorithms the amount of training data a child received through his or her daily experiences in both quantity and quality.

Once computers have visual intelligence that is the same as or better than that of humans, the positive applications are vast.

This was what ImageNet did starting in 2007. The project was the perfect undertaking for the confluence of forces mentioned previously: convolutional neural networks, which work best with large data sets; big data, or the hundreds of millions of digital images used in the project; and GPUs, which had advanced by this point to be able to process high-volume tasks efficiently.

Researchers employed crowdsourcing through Amazon's Mechanical Turk platform to identify and categorize nearly 1 billion images. At its peak, almost 49,000 people from 167 countries helped our researchers clean, sort, and label those candidate images. This represents the beginning of the third phase of object recognition.

Computer Vision Advances

If the goal of computer vision is to teach machines to see as humans do, where people are able to lay their eyes on something and nearly instantly weave entire stories and make meaning of people, places, and things, ImageNet represented a foundational building block of this by greatly advancing a computer's ability to identify a very large catalog of images.

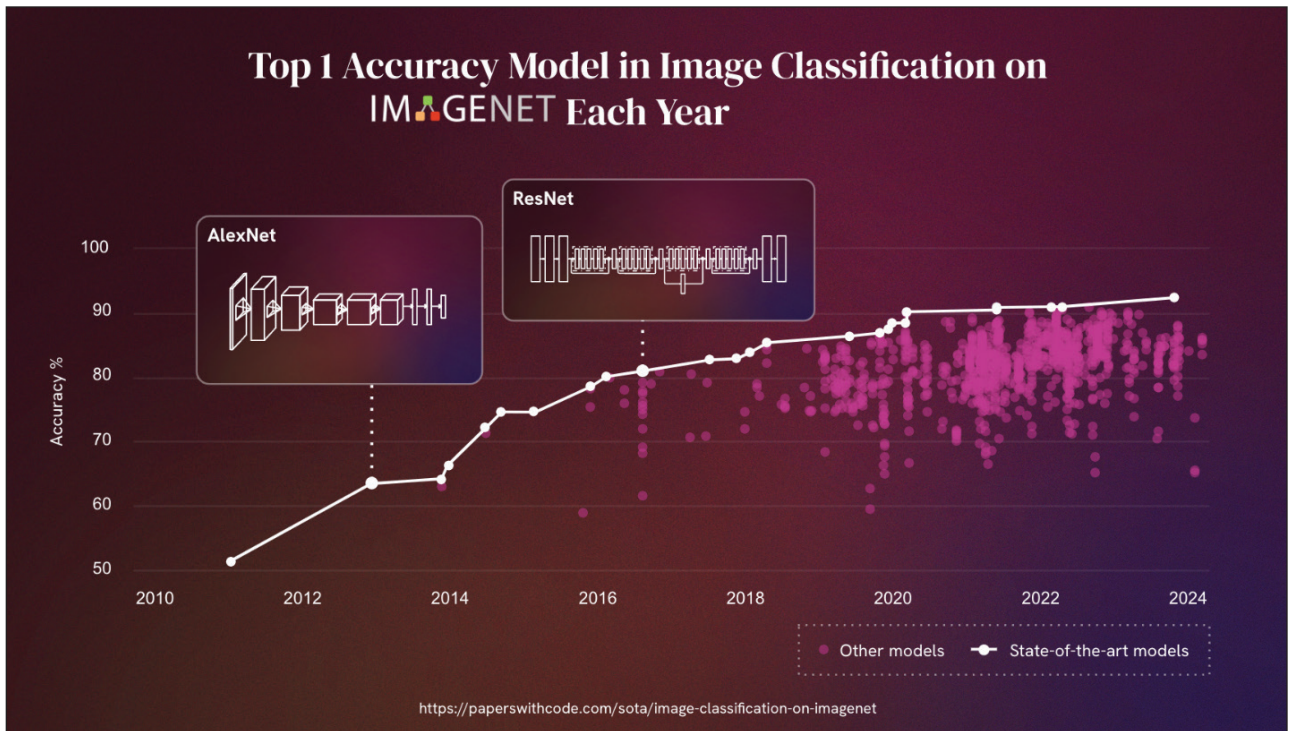


FIGURE 1. The historical performance of algorithms focused on image classification for the ImageNet annual challenge. It showcases the performance of leading methods, ranked by top-1 accuracy, and includes various architectures and advancements driving improvements in this benchmarking challenge.

By 2009, ImageNet represented a dataset of more than 15 million images across almost 22,000 categories of objects (Deng et al. 2009). Still, despite the scale of ImageNet, the result was simply teaching a computer to identify objects. To go back to the child analogy, this is like the child speaking a lot of nouns.

The next step was advancing to sentences, so rather than identifying a cat, understanding that, for example, the cat is lying on a bed or playing with a ball of yarn. To accomplish this, the connection between big data (the images) and the machine learning algorithms (the convolutional neural networks) processing them needed to evolve. In 2014, a group of researchers and I developed a model that was able to learn not only from the images but also from natural language sentences generated by people. The new models we designed worked like the human brain in that they integrated vision and language. They connected visual snippets of images with words and phrases (Karpathy and Li 2015).

This feat took a lot of hard work from a lot of people, but it still didn't match the level of context and meaning humans get from their vision. While the computer may accurately identify a picture of a boy and a birthday cake,

it still couldn't recognize that the boy was wearing a shirt given to him by his late grandfather or what type of cake it was. Nor could it tell you the emotion on the boy's face. These are the details that add so much meaning to what we see, so computer vision still had a long way to go.

The Era of Spatial Intelligence Is Upon Us

Over time, researchers designed faster, more accurate algorithms. Each year, my lab held the ImageNet Challenge, where we tested and tracked the progress of their algorithms. In the span of 13 years, the accuracy of these algorithms ballooned from just over 50% to just over 90% (see figure 1). This was significant, and now some of the latest algorithms can segment objects and predict the dynamic relationship between them in moving pictures (Gupta et al. 2023).

Since 2015, when we first got computers to describe a photo using natural language, there was a challenge to do that in reverse. In other words, rather than giving the computer the image and asking the computer to describe it, the challenge was to prompt the computer with what you want to see and ask it to generate something for you. This seemed impossible in 2015. Fast forward to today,

generative AI algorithms such as Sora by OpenAI, DALL-E, and many others, powered by a family of diffusion models that take human-prompted sentences and turn them into photos or videos, have achieved this capability.

And as progress with such large language model-driven generative algorithms races forward, in my research lab, my students and collaborators have developed a generative video model called Walt, which, while far from perfect, is creating some pretty compelling pieces. As impressive as this seems, there is still progress to be made because these models are still lacking the element of action. A computer that can see can do and learn, but it still cannot act upon the world.

This capability, to act upon what is seen, learned, and understood, is what I refer to as “spatial intelligence,” and it is the next frontier in computer vision. A

good example is figure 2. A computer can identify that there is a glass on a table, but it cannot do what our human brain does in an instant: understand the geometry of the glass, its place in 3-D space, its relationship with the table, cat, and everything else present, and instinctively want to act to stop the glass from tipping over.

Getting a computer to understand this is difficult and requires translating 2-D images into 3-D models. But this will be critical for computers to become spatially intelligent. This year, researchers in my lab were able to load a single 2-D image into a computer and use an algorithm that renders a 3-D image from it (Sargent et al. 2024).

At the University of Michigan, a group of researchers has figured out how to turn a descriptive sentence into a 3-D room layout (Höllein et al. 2023). At Stanford, researchers and their students have created an algorithm that takes a single image and generates infinite plausible spaces for viewers to explore (Ge et al. 2024).

These examples are the first signs of humans’ ability to model the richness and nuances of the 3-D world in digital forms. This rapid progression of spatial intelligence



FIGURE 2. An example of an image that illustrates spatial intelligence in humans. Our brains see this image and understand the relationship between the glass, the cat, and the table in 3-D space. One of the goals of spatial intelligence is to teach computers to understand this image.

is ushering in a new era of robotics by catalyzing robotic learning, something that is a key component for any embodied intelligence system that needs to understand and interact with the 3-D world.

Where ImageNet used high-quality photos to train computers to see, today researchers are training with behaviors and actions so robots can know how to act in the world. To make this possible, researchers are developing 3-D spatial environments powered by 3-D spatial models so the computers have infinite possibilities from which to learn (Ge et al. 2024).

There is also exciting progress happening in robotic language intelligence. Through large language model-based input, my students and our collaborators can input a sentence and have a robot perform an action such as opening a drawer, unplugging a phone charger, or even making a simple sandwich (Huang et al. 2023).

Real-World Impact of Spatial Intelligence

Once computers have visual intelligence that is the same as or better than that of humans, the positive applica-

tions are vast. Healthcare is one industry with many applications. Researchers from my Computer Vision Lab at Stanford University have been collaborating with Stanford's School of Medicine and hospitals to pilot smart sensors that can detect whether clinicians have entered a patient room without properly washing their hands.

AI can also aid with tasks such as keeping track of surgical instruments and alerting care teams if a patient is at risk of a fall. Imagine an autonomous robot transporting medical supplies while caretakers focus on patients, or augmented reality that guides surgeons to do safer, faster, and less invasive procedures.

With spatially intelligent AI, all of this and so much more is possible. What took 540 million years to evolve in humans will happen in computers in a matter of decades. And the human species will ultimately be the beneficiary.

References

- Biederman I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 94(2):115–47.
- Deng J, Dong W, Socher R, Li L-J, Li K, Li F-F. 2009. ImageNet: A large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. IEEE.
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* 392:598–601.
- Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A. 2010. The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision* 88:303–38.
- Ge Y, Tang Y, Xu J, Gokmen C, Li C, Ai W, Martinez BJ, Aydin A, Anvari M, Chakravarthy A, and 13 others. 2024. BEHAVIOR vision suite: Customizable dataset generation via simulation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22401–12. IEEE.
- Gupta A, Wu J, Deng J, Li F-F. 2023. Siamese masked autoencoders. *arXiv:2305.14344*.
- Höllein L, Cao A, Owens A, Johnson J, Nießner M. 2023. Text2Room: Extracting textured 3D meshes from 2D text-to-image models. *arXiv:2303.11989*.
- Huang W, Wang C, Zhang R, Li Y, Wu J, Li F-F. 2023. VoxPoser: Composable 3D value maps for robotic manipulation with language models. *arXiv:2307.05973*.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17(11):4302–11.
- Karpathy A, Li F-F. 2015. Deep visual-semantic alignments for generating image descriptions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3128–37. IEEE.
- Sargent K, Li Z, Shah T, Herrmann C, Yu H-X, Zhang Y, Chan ER, Lagun D, Li F-F, Sun D, and 1 other. 2024. ZeroNVS: Zero-shot 360-degree view synthesis from a single image. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9420–29. IEEE.
- Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381(6582):520–2.

AI innovations, developed and deployed responsibly, can help preserve our climate, improve health outcomes, and create a more accessible world for everyone.

AI's Capabilities Make It a Powerful Tool for Driving Societal Impact

Yossi Matias, Avinatan Hassidim,
and Philip Nelson



Yossi Matias



Avinatan Hassidim



Philip Nelson

Many of us entered the sciences driven by a profound curiosity and by a desire to solve the challenges around us and make new discoveries about our world. With artificial intelligence (AI), we are now living in a golden age of research, where the pace of innovation rivals the early days of the internet. The incredible capabilities of AI, including generative, predictive, and diagnostic, make it a transformative tool for addressing some of humanity's most complex challenges. We are already seeing how AI can help preserve our climate, improve health outcomes, propel scientific understanding, democratize education, and create a more accessible world for everyone.

Along with the significant benefits to humanity, there are also legitimate concerns about this nascent technology, from hallucinations and perpetuating

Yossi Matias is vice president at Google and the head of Google Research. Avinatan Hassidim is a vice president at Google Research. Philip Nelson is a director of engineering at Google Research.

biases to privacy and security risks. The development and deployment of AI technologies in a responsible and collaborative manner can help mitigate the risks and maximize the benefits, with alignment across stakeholders and the public and private sectors. Below, we demonstrate the critical role that AI can play in society, drawing on examples of AI innovations from teams at Google Research and collaborative efforts across the research ecosystem to ensure that AI, used responsibly, can truly benefit everyone.

Developing AI Solutions to Address Climate Crises

The climate crisis is one of today's most pressing global challenges; it demands urgent and decisive action. At forums such as the annual COP conference,¹ scientists and governments have aligned on the importance of mitigating the extent of climate change while simultaneously helping communities adapt to its catastrophic impact. AI is playing a crucial role in accelerating climate action. It is used to enhance our holistic understanding of the climate system, help mitigate the climate impact of various industries, improve our ability to predict and adapt to climate-related events, and optimize climate action strategies.

***Healthcare is another domain
where AI has already started
to deliver on its potential
to improve lives.***

Mitigating Climate Change

Take transportation, which is responsible for a significant amount of global and urban greenhouse gas emissions. Road transportation is especially problematic at city intersections, where pollution can be 29 times higher than on open roads (Kumar and Goel 2016). To help address this, Google Research's Project Green Light² uses AI-powered traffic flow modeling and Google Maps' comprehensive understanding of global road networks to create smart recommendations for traffic light optimization. Results from the pilots with cities indicate a potential for up to

a 30% reduction in vehicle stops and a 10% decrease in greenhouse gas emissions. Green Light is now being scaled and deployed around the world from Seattle to Bangalore to Jakarta.

In aviation, about 35% of the global warming impact comes from the clouds created by contrails, the condensation trails sometimes seen behind airplanes (Jaramillo 2022), which form in humid conditions when water vapor condenses around the tiny particles of pollutants emitted by airplane engines. By combining weather data, satellite data, and flight data, AI can predict when and where contrails are likely to form, enabling pilots to adjust the altitudes of their flights accordingly. American Airlines and Google Research demonstrated how airlines can reduce their climate impact: A program involving 70 flights over six months resulted in a 54% reduction in contrails compared to flights without AI-guided route adjustments (Elkin and Sanekommu 2023). At scale, such initiatives could have a meaningful, cumulative environmental impact.

While these AI-based solutions do not require high amounts of energy, it is important to note that AI technologies, particularly generative AI, often consume significant amounts of energy. This is a distinct challenge that merits its own mitigation solutions, such as more energy efficient models and hardware.

Helping Communities Adapt and Build Resilience

AI is being utilized to help communities build resilience and better prepare for and respond to natural disasters. It can analyze vast datasets, including historical records and real-time weather and satellite data, to predict and track extreme weather events with greater accuracy than previously possible. Forecasting floods is one such example. Every year, floods disrupt the lives of millions worldwide, causing thousands of fatalities and significant financial damages. Google Research's AI-powered forecasting system uses a state-of-the-art hydrological model to forecast the amount of water flowing in a river and is able to leverage global, publicly available data to generate predictions even in regions with limited historical data. This improves forecasts across regions in Africa and Asia to be similar to those that are currently available in Europe and provides coverage in 100 countries for 700 million people (Cohen 2024; Nearing et al. 2024), with flood forecasts distributed on Google Maps and Android devices. To maximize the value of the research, the data has been made available to researchers and partners through an application programming interface and through the Google Runoff

¹ www.un.org/en/climatechange/cop28

² <https://sites.research.google/greenlight/>



FIGURE 1 FireSat is the first satellite constellation to focus on early detection in high resolution imagery. Source: Google Research website (<https://sites.research.google/gr/wildfires/firesat/>).

Reanalysis & Reforecast dataset³ which includes forecasts dating back to 1981. Giving the ecosystem access to more and higher-quality data should enable the development of more effective mitigation strategies.

As global temperatures rise, wildfires are also devastating communities around the world, as seen recently in Los Angeles. Since 2020, Google Research has employed AI models to track wildfire boundaries and has surfaced alerts providing helpful information to people who are caught near a wildfire (Royz and Tendler 2024). This is now available in over 20 countries. To further advance wildfire research, Google Research released FireBench,⁴ an open-source machine learning (ML) benchmark dataset. Plus, Google Research collaborated with the Earth Fire Alliance and the Moore Foundation to contribute to the creation of FireSat (Van Arsdale 2024), a purpose-built constellation of satellites designed specifically to detect and track wildfires as small as a classroom (roughly 5x5 meters; figure 1). With FireSat, authorities will have near real-time information about the location, size, and inten-

sity of early-stage wildfires updated every 20 minutes so firefighters and emergency responders can respond effectively. FireSat's data will also be used to create a global historical record of fire spread, helping scientists to better model and understand wildfire behavior.

Advancing AI to Help More People Worldwide Live Healthier Lives

Healthcare is another domain where AI has already started to deliver on its potential to improve lives. It is being used to expand access to quality care worldwide, provide customized information to people on their health journeys, generate insights for public health authorities, and empower clinicians to provide timely, accurate diagnoses to patients. Going forward, AI-powered technologies will allow experts to reimagine patient care and will increasingly be used by both patients and clinicians, helping people everywhere live longer and healthier lives.

Improving Healthcare for Everyone Through AI Diagnostic Tools

AI accelerates medical imaging and diagnostics by rapidly analyzing vast amounts of complex data and providing objective assessments, which can lead to more timely and

³ The Google Runoff Reanalysis & Reforecast dataset is available at sites.research.google/gr/floodforecasting/resources/

⁴ sites.research.google/gr/wildfires/firebench/

effective interventions, and thus better health outcomes. Breast cancer, one of the most common cancers globally, is an example where early detection is possible through screenings and can lead to better chances of survival. Together with partners including iCAD (Corrado 2022), NHS, Imperial College London, and Northwestern Medicine, Google Research has developed and integrated AI models for mammography into breast cancer screening workflows to help radiologists identify breast cancer earlier and more consistently. Published research shows that this technology can identify signs of breast cancer as effectively as trained radiologists (McKinney et al. 2020). Google Research also developed ML tools for lung cancer screenings: The system analyzes CT scans and provides a cancer suspicion rating, helping radiologists identify potential malignancies more accurately (Kiraly and Pilgrim 2024). Study results indicate the potential for one person out of every 15–20 patients screened to be able to avoid unnecessary follow-up procedures, reducing their anxiety and the burden on the healthcare system (Kiraly et al. 2024). DeepHealth and Apollo Radiology International are exploring how to integrate this technology into clinical practice.

In a world with a shortage of teachers and overcrowded classrooms, there is a huge opportunity for AI to make a difference in formal education settings, helping both educators and students.


It is important to ensure that AI is also deployed in less economically developed countries, taking into account practical challenges such as unreliable internet connections, the availability of trained staff, and the need to operate within established frameworks for follow-up care. AI-powered diagnostic tools can play a crucial role in boosting access to care in low-resourced regions across the Global South. One example is screening for diabetic retinopathy, the leading preventable cause of blindness.

Over 500 million adults worldwide have diabetes,⁵ but many do not have access to eye specialists and the regular screenings required for the timely detection and treatment of the disease. Back in 2015, Google Research sought to address this together with the Aravind Eye Hospital in India and Rajavithi Hospital in Thailand, along with the founder of Thailand's national diabetic retinopathy screening program, Dr. Paisan Ruamviboonsuk. They pioneered Automated Retinal Disease Assessment, an AI-based diagnostic tool for high-quality diabetic retinopathy screening. Ophthalmologists helped researchers train the model, which can be operated by nurses and has screened over 600,000 patients to date. Forus Health, AuroLab, and Perceptra aim to reach six million people in India and Thailand over the next decade using this technology (Sawhney 2024). It is an example of how years of collaborative research can be piloted and then applied at scale in clinical settings to make quality care accessible to underserved populations.

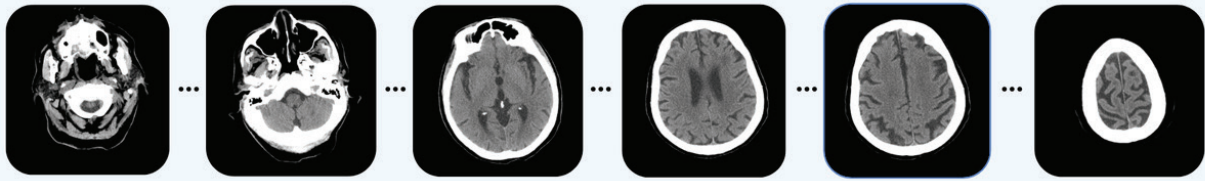
When developing AI applications for healthcare, it's important to prioritize patient safety and privacy, and to deliver equitable results regardless of the socioeconomic status, race, gender, or other demographic factors of patients. That is why piloting research in real-world settings, under medical guidance from partners, and with people of diverse backgrounds matters. Frameworks can also help prevent biases in AI datasets and tools. Last year, Google Research introduced the Health Equity Assessment of ML performance (Schaekermann et al. 2024), a four-step process for quantitatively estimating the performance of ML tools for groups with, on average, worse health outcomes. This could inform improvements in product development and real-world testing to make health AI technologies more effective for more people. The framework was tested on the Skin Condition Image Network dataset,⁶ developed by Google Research and Stanford University. It contains over 10,000 user-contributed images of skin, nail, or hair conditions, and aims to provide a more representative dermatology dataset across diverse skin tones and body parts. The intention is that this dataset will be a useful resource for all those working to advance inclusive dermatology research, education, and AI tool development. It is important for such datasets to be diverse and accurately reflect the complexities and variations within the population, so that when AI algorithms are trained on the data, they are free from biases and can promote better health outcomes for all.


⁵ <https://idf.org/about-diabetes/diabetes-facts-figures/>

⁶ <https://github.com/google-research-datasets/scin>


 **Prompt**

Given the image, what are the radiology findings?



 **Response**

FINDINGS: mild cortical atrophy and periventricular white matter changes are noted compatible with microvasculopathy. there is proportionate ventriculomegaly. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect. no abnormal extra-axial fluid collection is identified. no definitive evidence of intracranial hemorrhage or large territorial infarct within the limits of ct imaging. **old left basilar ganglia infarct**. no intracranial mass, swelling, hemorrhage, infarction or extra-axial collection is seen. visualized mastoid air cells: well pneumatized. visualized paranasal sinuses: well pneumatized. atherosclerotic calcification of vasculature at the skull base.

 **Radiologist Written Report**

FINDINGS: the fourth ventricle is midline. the remaining ventricular system is within normal limits. there is mild chronic white matter microangiopathic disease and moderate generalized cortical atrophy. no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. bone windows demonstrate an intact skull. the remaining visualized paranasal sinuses and mastoids are clear. chronic white matter microangiopathic disease, mild generalized cortical atrophy, not significantly changed since 2015. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.

FIGURE 2 Med-Gemini-3D is able to generate reports for CT scans, a far more complex form of imaging than standard X-rays. In this example, Med-Gemini-3D's report has correctly included a pathology (highlighted in green) that was missed in the original radiologist's report. Note that 'basilar' is a common mistranscription of 'basal' that Med-Gemini has learned from the training data, though the meaning of the report is unchanged. Source: Google Research blog (<https://research.google/blog/advancing-medical-ai-with-med-gemini/>).

Delivering Personalized Healthcare with Generative AI

Arguably the most profound shift to date in AI research has been the recent and rapid rise of generative AI and its mainstream adoption across society. Now being applied to the medical domain, this technology is poised to enable highly personalized insights which could revolutionize the care patients receive and pave the way for more preventative healthcare solutions.

Google Research is exploring the value of generative AI in healthcare to assist clinicians, researchers, and patients. Notably, Med-Gemini (Corrado and Barral 2024) is a family of AI models fine-tuned for medical applications. On the MedQA benchmark, which uses US Medical Licensing Exam-style questions, Med-Gemini achieved a state-of-the-art accuracy of 91.1%, surpassing previous models (Saab et al. 2024). It can interpret complex 3D scans, answer clinical questions, and generate state-of-the-art radiology reports (figure 2). These results are a testament to the advanced multimodal and reasoning capabilities of today's generative AI and lay the foundation for more accurate and individualized medical care.

For all the advances in AI technology, the physician-patient conversation is, and will remain, a cornerstone of medicine. Google Research is conducting research to inform how generative AI could be used to augment diagnostic medical reasoning and conversations between clinicians and patients. AMIE (Articulate Medical Intelligence Explorer; Karthikesalingam and Natarajan 2024) is a large language model (LLM)-based research system trained and evaluated along many dimensions that reflect quality in real-world clinical consultations from the perspectives of both clinicians and patients (Tu et al. 2024). In simulated consultations, AMIE performed at least as well as primary care physicians when evaluated on clinically meaningful aspects of consultation quality, and clinicians assisted by AMIE arrived at more comprehensive differential lists than those without AMIE assistance. Following these promising results, AMIE is being tested with healthcare organizations to see how it could support clinical conversations, with oversight from medical professionals.

As a powerful tool for delivering insights, AI could also help personalize and improve overall wellness.

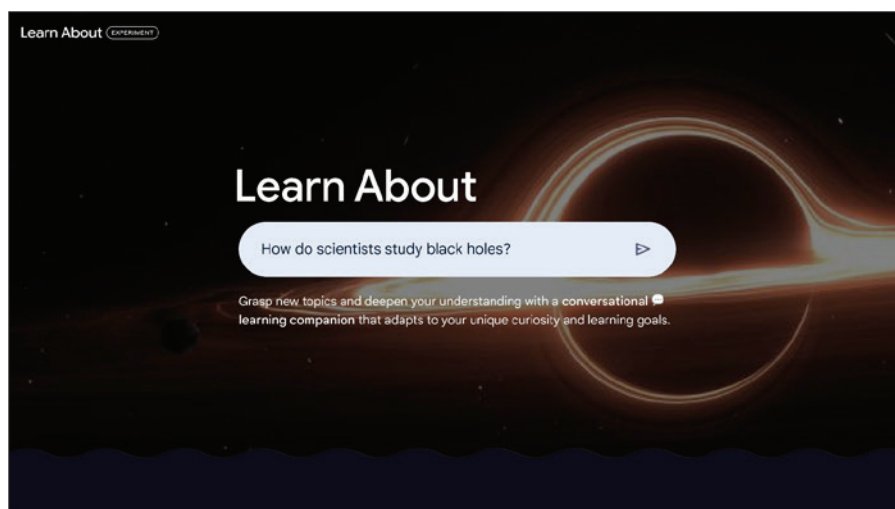


FIGURE 3 Learn About is a new experiential product that explores how information can turn into understanding by bringing together high-quality content, learning science, and chat experiences. Users can ask a question, and it helps guide them through any topic at their own pace, with pictures, videos, web pages, and activities, and allows them to ask clarifying questions. Source: Google Keyword blog (<https://blog.google/outreach-initiatives/education/google-learnlm-gemini-generative-ai/>).

Google Research is exploring this with projects such as the Personal Health Large Language Model (PH-LLM) and LLM agents that analyze wearable data from Fitbits and other devices (Merrill et al. 2024). The PH-LLM, a fine-tuned version of Google's Gemini model, is designed to give personalized advice on sleep and fitness. It is trained to understand both written text and data from wearable sensors. In tests, PH-LLM's fitness recommendations were as good as those from human experts. The model demonstrated a strong understanding of health topics, scoring 79% on sleep-related and 88% on fitness multiple-choice tests, outperforming average human scores (Cosentino et al. 2024). This research shows the significant progress and potential of AI models to provide unique health insights to users. Given the sensitivities around personalized healthcare and insights, it is important to undertake this research with strong attention to user privacy and safety.

Empowering the Next Generation of Learners with AI Tools Grounded in Learning Science

Learning is limitless. People learn as students, on the job, for a new career, and throughout their lives, pursuing hobbies and becoming curious about new things as their lives change and interests evolve. Generative AI is unlocking the potential within learners of all ages by democratizing access to information, personalizing

experiences, and encouraging creativity. It now allows for engaging conversational experiences that foster understanding and knowledge acquisition.

In collaboration with pedagogy experts, Google Research developed LearnLM,⁷ a family of AI models fine-tuned for education. LearnLM is grounded in educational research and tailored to how students learn, with the aim of creating more engaging, personalized, and effective learning experiences (Gomes 2024). In a recent technical report, LearnLM outperformed other leading AI models when it comes to adhering to the principles of learning science (Wiltberger 2024), such as explaining concepts at appropriate levels, providing effective

guidance, interactively guiding learners, and encouraging active engagement. The technology is already integrated into existing platforms, including Google Search and YouTube,⁸ to enhance understanding rather than simply providing answers. For example, in Google Search, AI Overviews can be adjusted to simplify language or break down complex topics.

Nothing can replace the magic of the teacher-student relationship. But in a world with a shortage of teachers and overcrowded classrooms, there is a huge opportunity for AI to make a difference in formal education settings, helping both educators and students. LearnLM is being used to create new educational tools, such as Illuminate, an experiment that breaks down research papers into audio conversations, and Learn About, an experiential tool that guides learners through various topics using different media and activities (figure 3). AI could also serve as a personal assistant to teachers, supporting them with tasks such as lesson planning, testing, and grading, enabling them to spend more time with their students. Google is piloting programs in Google Classroom⁹ to simplify lesson planning, for example, suggesting multimedia

⁷ https://services.google.com/fh/files/misc/improving-gemini-for-education_v7.pdf

⁸ <https://www.youtube.com/@Googlehelp>

⁹ <https://edu.google.com/workspace-for-education/products/classroom/>

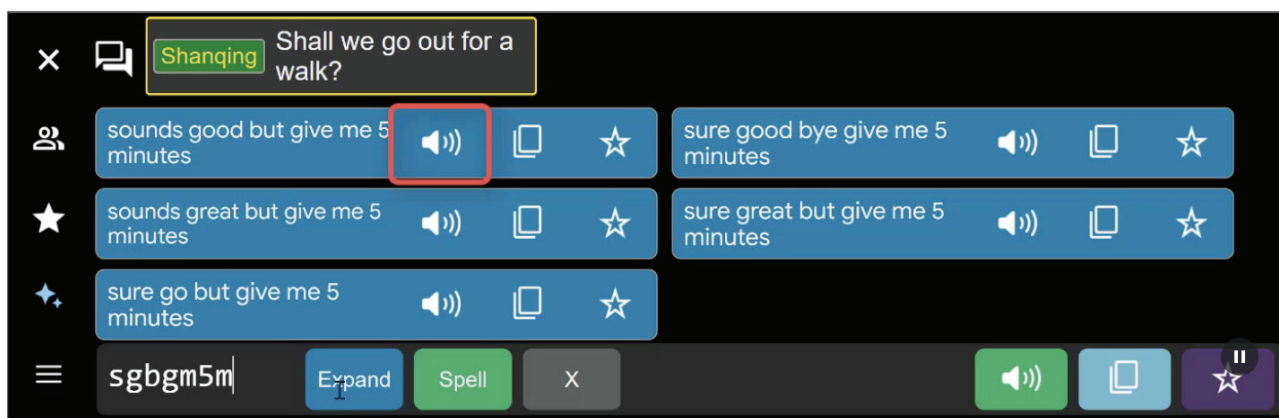


FIGURE 4 Screencast of the SpeakFaster user interface in action. The interface allows users to enter abbreviated text as input and uses the context of the conversation to suggest responses in full sentence form. Source: Google Research blog (<https://research.google/blog/speakfaster-revolutionizing-communication-for-people-with-severe-motor-impairments/>).

resources about a topic or allowing educators to adjust their content for different class levels.

Given the access young students will have to these tools, it is vital to emphasize safety, data privacy, and age-appropriateness. Tech companies have a responsibility to consult with experts on child safety and development, partner with pedagogical experts, conduct rigorous testing, and provide resources for responsible AI use.

Employing AI to Create a More Accessible World for Everyone

Groups who may be at risk of being overlooked are sometimes the ones who stand to benefit the most from AI. AI-powered technologies can help create more inclusive, accessible, and personalized experiences for people with disabilities in their everyday lives. By better understanding user interactions and adapting interfaces in real time, AI can give users with disabilities access to information and communication tools so they can interact more easily with other people and with the digital world.

For people who are d/Deaf or hard of hearing, AI can provide real-time transcriptions of speech and other sounds, making daily conversations and video and audio content more accessible. This technology is seen in products like Google's Live Transcribe, which provides real-time speech-to-text transcriptions, and Live Caption, which automatically generates captions on videos and audio across Android devices and Chrome browsers. For people who are blind or low-vision, AI can read online content out loud, providing the most relevant information first and offering descriptions of images, making it more helpful than traditional screen readers. The newly released NotebookLM can convert written documents

or long videos into an engaging conversation about the topic, which can benefit many users.

AI can also help the hundreds of millions of people globally who have non-standard speech to be heard and understood. Project Relate,¹⁰ for example, is an Android app developed by Google Research that uses machine learning to understand an individual's non-standard speech patterns. It enables users to communicate more easily through features such as real-time transcription, synthesized voice repetition, and direct interaction with Google Assistant.¹¹ AI can also help people with conditions like amyotrophic lateral sclerosis (ALS) and similar motor impairments to communicate through eye movements. These users often rely on speech-generating technology, but it can be extremely slow and therefore limiting. Google Research partnered with Team Gleason,¹² a non-profit organization that serves individuals with ALS, to create SpeakFaster (figure 4), a prototype that uses large language models to accelerate eye-gaze-based typing by predicting full phrases from abbreviated text and conversational context with high accuracy. Initial studies demonstrate that the system requires 57% fewer motor actions than traditional predictive keyboards in offline simulations, resulting in text-entry rates 29-60% faster than traditional baselines (Cai et. al. 2024). In addition, AI speech synthesis can be used to recreate voices for people with ALS, allowing them to communicate with friends and loved ones in a more authentic manner than standard, robotic-sounding voices of traditional

¹⁰ <https://sites.research.google/relate/>

¹¹ <https://assistant.google.com/>

¹² <https://teangleason.org/>

solutions. Such technologies can help remove barriers to the world and meaningfully change lives.

Removing language barriers can also help make the world more accessible. Much of the investment in language models to date has been in English. However, multilingual AI can help break down language barriers and help make online information and other AI tools accessible to users from all backgrounds. Google Assistant can now read web articles aloud in 42 languages and can translate the text when needed, and Google Translate recently expanded to include 110 new languages using AI, increasing access to information for over 614 million more people (Caswell 2024). It is also valuable to invest efforts in multilinguality so that language models are accurate and accessible in different languages and are relevant to people across different countries and cultures.

AI Must Be Deployed Responsibly

For all the recent progress, the use of AI to advance societal causes is still in the early stages. As the technology matures, society will benefit from a focus on safety for children, patients, and less advantaged populations, from an emphasis on both quality and equality, and from putting checks and regulations in place. If the risks and challenges can be properly addressed, there will be a plethora of opportunities to create useful, real-world applications that optimize classrooms, clinical settings, urban spaces, online access, and more.

Google's approach is grounded in AI principles that guide its efforts to develop the technology responsibly and collaboratively so that it works for everyone. Google advocates for bold innovations that can advance the frontier of AI and scientific research to benefit people everywhere, and the organization develops and deploys AI innovations responsibly, in a way that addresses user needs and rights such as safety, security, and privacy. That includes implementing appropriate due diligence and feedback mechanisms, aligning with international legal principles, and employing rigorous design, testing, and monitoring to mitigate any harmful outcomes and avoid unfair bias. This work is best done in collaboration with a wide range of partners, including researchers across industry and academia and with governments and civil society, to maximize the benefits of AI for all.

AI, developed and deployed boldly, responsibly, and collaboratively, can help provide solutions to previously unsolvable problems. It is already having a profound impact, and as we look to the future, the potential of this technology to address societal challenges and improve

people's lives has never been clearer. Stakeholders from the public and private sectors have good reason to come together to realize the vast AI opportunity and establish partnerships to scale the technology safely in domains like healthcare and education. The authors of this article look forward to joining forces with others dedicated to leveraging AI to benefit humanity as we ask the biggest of questions and pioneer groundbreaking answers. We are committed to pursuing AI responsibly, and with the conviction and determination that it will be a powerful force for good.

References

- Cai S, Venugopalan S, Seaver K, Xiao X, Tomanek K, Jalasutram S, Ringel Morris M, Kane S, Narayanan A, MacDonald RL, and 6 others. 2024. Using large language models to accelerate communication for eye gaze typing users with ALS. *Nature Communications* 15:9449.
- Caswell I. 2024. 110 new languages are coming to Google Translate. Google, The Keyword, June 27.
- Cohen D. 2024. An improved flood forecasting AI model, trained and evaluated globally. Google Research, Nov 11.
- Corrado G. 2022. Partnering with iCAD to improve breast cancer screening. Google, The Keyword, Nov 28.
- Corrado G, Barral J. 2024. Advancing medical AI with Med-Gemini. Google Research, May 15.
- Cosentino J, Belyaeva A, Liu X, Furlotte N, Yang Z, Lee C, Schenck E, Patel Y, Cui J, Douglas Schneider L, and 24 others. 2024. Towards a Personal Health Large Language Model. arXiv:2406.06474.
- Elkin C, Sanekommu D. 2023. How AI is helping airlines mitigate the climate impact of contrails. Google, The Keyword, Aug 8.
- Gomes B. 2024. How generative AI expands curiosity and understanding with LearnLM. Google, The Keyword, May 14.
- Jaramillo P, Ribeiro SK, Newman P, Dhar S, Diemuodeke OE, Kajino T, Lee DS, Nugroho SB, Ou X, Hammer A, and 2 others. 2022. Transport. In: IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Shukla PR, Skea J, Slade R, Al Khourdajie A, van Diemen R, McCollum D, Pathak M, Some S, Vyas P, Fradera R, Belkacemi M, Hasija A, Lisboa G, Luz S, Malley J, eds. Cambridge University Press.
- Karthikesalingam A, Natarajan V. 2024. AMIE: A research AI system for diagnostic medical reasoning and conversations. Google Research, Jan 12.

- Kiraly AP, Cunningham CA, Najafi R, Nabulsi Z, Yang J, Lau C, Ledsam JR, Ye W, Adrila D, McKinney SM, and 15 others. 2024. Assistive AI in lung cancer screening: A retrospective multinational study in the United States and Japan. *Radiology: Artificial Intelligence* 6(3):e230079.
- Kiraly A, Pilgrim R. 2024. Computer-aided diagnosis for lung cancer screening. Google Research, March 20.
- Kumar P, Goel A. 2016. Concentration dynamics of coarse and fine particulate matter at and around signalised traffic intersections. *Environmental Science: Processes & Impacts* 18:1220–35.
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, and 21 others. 2020. International evaluation of an AI system for breast cancer screening. *Nature* 577:89–94.
- Merrill MA, Paruchuri A, Rezaei N, Kovacs G, Perez J, Liu Y, Schenck E, Hammerquist N, Sunshine J, Tailor S, and 10 others. 2024. Transforming wearable data into health insights using large language model agents. *arXiv:2406.06464*.
- Nearing G, Cohen D, Dube V, Gauch M, Gilon O, Harrigan S, Hassidim A, Klotz D, Kratzert F, Metzger A, and 8 others. 2024. Global prediction of extreme floods in ungauged watersheds. *Nature* 627:559–63.
- Royz M, Tendler A. 2024. Wildfire boundary maps expand to new countries in Europe and Africa. Google, The Keyword, July 29.
- Saab K, Tu T, Weng WH, Tanno R, Stutz D, Wulczyn E, Zhang F, Strother T, Park C, Vedadi E, and 57 others. 2024. Capabilities of gemini models in medicine. *arXiv:2404.18416*.
- Sawhney R. 2024. How AI is making eyesight-saving care more accessible in resource-constrained settings. Google Research, Oct 17.
- Schaekermann M, Spitz T, Pyles M, Cole-Lewis H, Wulczyn E, Pfohl S, Martin Jr D, Jaroensri R, Keeling G, Liu Y, and 17 others. 2024. Health equity assessment of machine learning performance (HEAL): a framework and dermatology AI model case study. *The Lancet eClinicalMedicine* 70:102479.
- Tu T, Palepu A, Schaekermann M, Saab K, Freyberg J, Tanno R, Wang A, Li B, Amin M, Tomasev N, and 15 others., 2024. Towards conversational diagnostic AI. *arXiv:2401.05654*.
- Van Arsdale C. 2024. A breakthrough in wildfire detection: How a new constellation of satellites can detect smaller wildfires earlier. Google, The Keyword, Sept 16.
- Wiltberger S. 2024. LearnLM outperformed other AI models in a recent technical study. Google, The Keyword, Dec 19.

Governance should not be understood as an impediment to AI innovation but as an essential component of it.

Disrupting the Disruption Narrative: Policy Innovation in AI Governance

Alondra Nelson



Alondra Nelson (NAM) is the Harold F. Linder Professor at the Institute for Advanced Study.

“Disrupt!” has been a mantra of technology-driven commerce for more than three decades. Clayton M. Christensen, the late Harvard Business School professor, pioneered the analysis of this idea in practice—which he termed “disruptive innovation”—developing influential research that highlighted strategies for identifying novel approaches to capture incumbent business and emerging markets (1997).

Disruption of the existing market economy manifests in many ways. It can be technological, such as in the late 20th century when vacuum tubes were overtaken by transistors, revolutionizing semiconductor development and consumer electronics production by enabling the creation of products that were inexpensive, efficient, and portable (Riordan and Hoddeson 1998). Further-

more, disruption occurs by reshaping markets through the introduction of products that may be more accessible and affordable—often marketed as “free” despite hidden costs—and thereby expanding the consumer base (Christensen 1997; Shapiro and Varian 1998; Terranova 2000).

One prevalent depiction of the disruption economy appears as a foot race, with the “hare” of technological innovation speeding past the slow-moving “tortoise” of policy and governance. This characterization has gained particular traction in discussions about artificial intelligence (AI) as its expanding use across society has elevated the matter of technology governance to one of the most pressing challenges of our time. The accompanying narrative that agile, meaningful oversight is impossible due to the speed of innovation has become especially entrenched in discussions of AI systems, which are said to evolve so rapidly and to transform society so fundamentally that policy frameworks cannot possibly keep pace.

This prevailing perspective on the relationship between disruption and innovation is both incomplete and inaccurate. A critical missing element is the recognition that some tech industry actors purposely seek to evade or resist regulatory frameworks as a deliberate business strategy (Edelman and Geradin 2016), integrating into their core mission not only technological and consumer-facing transformations but also regulatory arbitrage (Cohen 2019; Zuboff 2019). Moreover, this approach includes tactics for skirting, bending, circumventing, or resisting existing legal frameworks (Hussain et al. 2020; Rahman and Thelen 2019), and serves as its own engine of disruption,

not merely a secondary effect or unintended consequence of it.

A Cautionary Tale for AI Governance: Ridesharing Companies and Regulatory Arbitrage

The rise of ridesharing fundamentally transformed mobility patterns, traffic congestion, labor relations, and public transportation ecosystems (Calo and Rosenblat 2017; Shaheen et al. 2016), causing a range of harms from the erosion of workers’ rights and quality of life (Dubal 2017; Malin and Chandler 2017) to increased traffic congestion and associated pollution risks (Erhardt et al. 2019) to consumers’ personal safety and discrimination (Ge et al. 2020; Hoskins 2022) to surveillance of users (Rosenblat 2018). Ridesharing companies have systematically shaped public discourse around their negative impacts by employing strategic regulatory avoidance and undermining existing legal frameworks (Rahman and Thelen 2019). They divert attention from their role in creating harms to workers, transportation systems, and urban infrastructure, presenting themselves as innovators, while actively subverting regulatory oversight designed to protect public interests. Ridesharing companies’ efforts to set the terms of debate on the public harms to which they contribute exemplify a strategy of deliberate regulatory evasion and defiance. The ridesharing example illustrates key considerations for AI policymaking as AI companies similarly seek to set the conditions of their own governance by disregarding or undermining rules, laws, and policies, with concomitant harms.

When the ridesharing company Uber emerged in the United States, it strategically entered markets where the regulation of alternative transportation services remained undefined and unlegislated (Christensen et al. 2015; Rahman and Thelen 2019). A decade ago, Uber and Lyft drivers in Utah faced tickets and substantial fines on behalf of the companies until they met statewide requirements for background checks, liability insurance, and other public safety benchmarks (Price 2015). In New York City, ridesharing startups confronted an established taxi medallion system, which provided policymakers with legal frameworks to temporarily resist Uber’s entry into one of the world’s largest transportation markets (Dubal 2017). Ridesharing companies intentionally operated outside existing taxi and transportation regulations, arguing that their technology-enabled services constituted an entirely new category requiring different legal treatment (Davis 2015; Thelen 2018).

Some tech industry actors purposely seek to evade or resist regulatory frameworks as a deliberate business strategy, integrating into their core mission not only technological and consumer-facing transformations but also regulatory arbitrage.

Understanding regulatory arbitrage as a calculated strategy (Pollman and Barry 2017) brings perceived gaps between innovation and regulation into clearer focus as deliberately engineered outcomes serving specific business interests, rather than as an inevitable consequence of the pace of technology (Christensen et al. 2015; Pasquale 2015; Zuboff 2019). When we view disruption as a strategy to circumvent laws, rather than as an inevitable outgrowth of technological development, we better understand how this approach proactively undermines regulatory guardrails (Cohen 2019). Moreover, fully understanding this dynamic opens up new possibilities for AI governance strategy by engaging in disruptive innovation *through policy innovation*.

Building on this understanding, we see how casting governance as a drag on innovation severely limits the spectrum of possibilities for effective AI policy, hampering the development and implementation of crucial organizational, corporate, and governmental guardrails needed to mitigate risks and prevent harm; ensure the safe design, production, and deployment of new technologies; and harness their potential. This restrictive framework fundamentally limits how policymakers and the public conceptualize and pursue viable approaches to AI governance. The commercial deployment of generative AI has precipitated numerous ongoing legal challenges concerning training data provenance (Grynbaum and Mac 2023), intellectual property rights (Brittain 2025), and competition law violations (Ciaccia 2024), collectively representing an emerging wave of governance disruption. But this new wave of disruption presents an opportunity to challenge the presumed inevitability of regulatory lag and to begin to leverage policy innovation to achieve more beneficial outcomes for AI use.

AI Policy Innovation: A Multi-Faceted Approach

Addressing these challenges requires a renewed commitment to *AI policy innovation*. Policy enables desired future states across institutions, while policy innovation encompasses the conditions necessary for achieving these desired states, including strategic development and implementation of novel approaches to principles, rules, and guidelines that can address governance challenges. The perception that AI governance inherently lags behind technological development overlooks an immediate solution: the application of existing laws, rules, regulations, and standards. A significant barrier to this approach, despite being the most agile response to emerging technol-

ogy, has been the persistent industry framing of AI—like many transformative technologies before it—as so fundamentally novel that existing governance frameworks cannot possibly address it (Selbst and Barocas 2018).

This narrative has led to AI being characterized as essentially ungovernable. The path to effective AI governance begins with demystifying artificial intelligence itself (Crawford 2021). While AI systems demonstrate remarkable and expanding capabilities, they remain fundamentally human-created tools with specific limitations and constraints. This foundational understanding helps to counter narratives that can paralyze effective policymaking. By recognizing AI as a product of human choices and decisions, we maintain a clearer perspective on our agency in shaping its development and deployment (Winner 2021). This demystification enables more pragmatic and effective governance approaches.

By recognizing AI as a product of human choices and decisions, we maintain a clearer perspective on our agency in shaping its development and deployment.

To strengthen AI governance despite industry resistance, policymakers can pursue three complementary approaches: First, they can more effectively leverage and maximize existing regulatory frameworks and legal mechanisms by adapting and applying current legal frameworks to AI challenges. Second, they can develop the new policy instruments and governance structures that may be required to address unique aspects of AI systems. And third, they can embrace an iterative approach to policy development that allows for rapid learning, adjustment, and evolution as technologies and their impacts continue to emerge. This multi-faceted approach to policy innovation enables more responsive and effective governance while avoiding the false choice between public protection and technological development.

Existing Laws

Guardrails are essential enablers, not obstacles. In response to criticisms that regulation is too slow, governments and organizations can resist industry rhetoric that new technologies like AI transcend all prior conceptions of laws, norms, and rules. Using this approach, safeguards can be swiftly deployed through the innovative application of existing governance frameworks—including established norms, rules, laws, and standards. While many of these laws were not designed with AI in mind, their intended outcomes—including safety, inclusion, accessibility, and equitable use—remain vital goals even as technology evolves.

Guardrails are essential enablers, not obstacles.

In some instances, leveraging existing laws and policies represents the most expedient path to responsive technology governance. For example, President Biden's 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence emphasized that "the use of new technologies, such as AI, does not excuse organizations from their legal obligations, and hard-won consumer protections are more important than ever in moments of technological change." The order stated that the "Federal Government will enforce existing consumer protection laws and principles and enact appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI" (EOP 2023). While executive orders may be subject to changes by subsequent administrations—this one was revoked by the Trump administration in early 2025—this example demonstrates the crucial principle that new technologies do not necessarily require reconstructing the social compact.

In practice, this governance approach has been manifested through concrete actions. The US Equal Employment Opportunity Commission issued guidelines applying the Americans with Disabilities Act to the use of software, algorithms, and AI in hiring practices (EEOC 2022). Similarly, the Federal Trade Commission launched "Operation AI Comply" to pursue cases against companies that used "AI tools to trick, mislead, or defraud people" (FTC 2024). Furthermore, state-level

legal and regulatory frameworks provide valuable models for AI governance. Illinois' Biometric Information Privacy Act, for instance, offers a template for protecting individual privacy rights in the AI era, while existing anti-discrimination laws provide mechanisms for addressing algorithmic bias (Citron and Pasquale 2014). In addition, federal agencies are actively developing guidance and rules to adapt conventional laws concerning intellectual property, copyright, and fair use to the AI context (e.g., USPTO 2024).

Effective AI governance must be grounded in fundamental democratic values and human rights. Just as the US Bill of Rights established essential protections for American democracy, AI governance frameworks must articulate and protect core societal values (Blueprint for an AI Bill of Rights, OSTP 2022); Lander and Nelson 2021). This requires balancing innovation with the public good, ensuring algorithmic systems respect human dignity and rights, and maintaining democratic oversight of increasingly powerful technologies.

In the courts, writers, artists, musicians, and media companies have filed copyright suits against AI companies (Brittain 2023), seeking to use existing law regarding the deployment of these tools and systems in creative industries. While the outcomes of these cases remain to be seen—and the plaintiffs could lose—these legal proceedings represent regulatory enforcement and also model innovative uses of law and policy for the AI era.

New Laws

The creation of innovative policy tools and governance frameworks is essential for addressing AI's distinctive challenges. Although conventional governance approaches remain useful, effective AI regulation may require developing diverse mechanisms across many institutional settings. We're already seeing policy expansion attempts in cases like *Mobley v. Workday*, where plaintiffs argue that Workday, an AI software vendor, should be classified as an "employer" under employment discrimination law (Wiessner 2024).

Concurrently, labor organizations have emerged as significant actors in AI governance, as demonstrated by the strategic actions of the Screen Actors Guild-American Federation of Television and Radio Artists, which successfully negotiated substantial concessions regarding AI implementation, including specific restrictions on the creation and utilization of both "digital replicas" of human performers and AI-generated "synthetic performers" (Franzen 2023).

State-level legislative initiatives further exemplify this multifaceted approach to AI governance. In California, although the widely debated Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (SB 1047) was vetoed by Governor Gavin Newsom, he signed into law 18 other AI-related bills in 2024, including measures to address election deepfakes and enhance protections for actors, which demonstrates the breadth of regulatory responses (Ables and De Vynck 2024). Similarly, the 2024 Colorado Artificial Intelligence Act establishes comprehensive obligations for AI developers and deployers, requiring them to protect consumers from foreseeable risks and harms, including algorithmic discrimination in crucial domains such as employment, education, housing, insurance, lending, and healthcare services (Colorado Consumer Protections for Artificial Intelligence Act 2024). These governance efforts illustrate how institutions are not merely responding reactively but are proactively and innovatively establishing meaningful parameters for AI development and deployment, particularly where it intersects with social and economic welfare.

Iteration

The dynamic nature of AI technology, however, presents unprecedented governance challenges that distinguish it from 20th-century technological innovations. Unlike relatively static technologies such as automobiles and semiconductors, some AI systems demonstrate the capacity for continuous evolution. This fundamental characteristic necessitates a reconceptualization of regulatory approaches, as traditional frameworks predicated on stable definitions and clear boundaries may prove insufficient for governing such dynamic systems.

The US Department of Commerce's National Institute for Standards and Technology (NIST) offers a model for an innovative approach to AI governance. Building upon its constitutional mandate to "fix the standard of weights and measures," NIST has expanded beyond its traditional role of establishing fundamental measurement standards to address the complexities of AI systems. Its AI Risk Management Framework 1.0 (2023) represents a significant departure from conventional standards development, introducing software development practices such as versioning into government standard-setting processes. While this framework—created collaboratively with industry, academia, and civil society—remains voluntary, effective AI governance requires a combination of laws, norms, and standards. Such adaptable mechanisms are essential components of a comprehensive approach.

This novel approach to AI policy acknowledges that contemporary technological systems extend far beyond basic measurement, encompassing complex decision-making capabilities that demand more sophisticated governance frameworks. The framework's sociotechnical orientation recognizes that effective standards must address not only technical specifications but also human and societal factors—specifically, how AI systems interact with and impact individuals and communities in real-world contexts. The integration of versioning practices, while commonplace in software development, represents a significant innovation in governmental standard-setting. This approach demonstrates how governance frameworks can be both robust and adaptable. The development of iterative guidelines, rules, and norms has become essential for effective governance of emerging technologies, particularly advanced AI systems.

Policy Innovation as Positive Disruption

The prevailing discourse around disruption has typically cast governance as an impediment to innovation. However, effective governance and technology development are not opposing forces but complementary elements in creating safe, sustainable, trustworthy, and beneficial AI systems. Policy innovation—whether through the application of existing frameworks, the development of new governance mechanisms, or the adoption of iterative approaches—represents its own form of positive disruption. This disruption manifests not as a circumvention of necessary guardrails but as a creative force that can catalyze technological ingenuity and protect and enhance societal wellbeing.

Effective governance and technology development are not opposing forces but complementary elements in creating safe, sustainable, trustworthy, and beneficial AI systems.

By recognizing governance as an essential component of technological development and deployment rather than an obstacle to it, we open new possibilities for addressing AI's challenges and opportunities. These examples of policy innovation—from NIST's versioned frameworks to state-level legislative initiatives—demonstrate that governance can be both robust and adaptable, creating a foundation for AI development that is both innovative and responsible. As we continue to navigate the complexities of AI governance, the understanding that true innovation involves constructive disruption that embraces socially responsible technology design and use will be essential.

Acknowledgments

Thank you to Chiraag Bains and Hannah Bloch-Wehba for their thoughtful, incisive feedback which helped to improve this essay.

References

- Ables K, De Vynck G. 2024. California passes AI laws to curb election deepfakes, protect actors. *The Washington Post*, Sept 18.
- Brittain B. 2025. Anthropic reaches deal in AI guardrails lawsuit over music lyrics. *Reuters*, Jan 3.
- Calo R, Rosenblat A. 2017. The taking economy: Uber, information, and power. *Columbia Law Review* 117(6):1623–90.
- Christensen CM. 1997. The innovator's dilemma: When new technologies cause great firms to fail. Harvard Business School Press.
- Christensen CM, Raynor M, McDonald R. 2015. What is disruptive innovation? *Harvard Business Review* 93(12):44–53.
- Ciaccia C. 2024. Google's partnership with Anthropic formally probed by UK. *Seeking Alpha*, Oct 24.
- Citron DK, Pasquale F. 2014. The scored society: Due process for automated predictions. *Washington Law Review* 89(1):1–33.
- Cohen JE. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press.
- Colorado Consumer Protections for Artificial Intelligence Act, SB24-205, 2024 Regular Session 2024. Online at <https://leg.colorado.gov/bills/sb24-205>.
- Crawford K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Dubal VB. 2017. The drive to precarity: A political history of work, regulation, and labor advocacy in San Francisco's taxi and Uber economies. *Berkeley Journal of Employment and Labor Law* 38(1):73–135.
- Edelman B, Geradin D. 2016. Efficiencies and regulatory short-cuts: How should we regulate companies like Airbnb and Uber? *Stanford Technology Law Review* 19:293–328.
- EEOC (US Equal Employment Opportunity Commission). 2022. The Americans with Disabilities Act and the use of software, algorithms, and artificial intelligence to assess job applicants and employees. Technical Assistance Document No. EEOC-NVTA-2022-2.
- Erhardt GD, Roy S, Cooper D, Sana B, Chen M, Castiglione J. 2019. Do transportation network companies decrease or increase congestion? *Science Advances* 5(5):eaau2670.
- EOP (Executive Office of the President). 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Executive Order 11410, 88 Fed. Reg. 75191, Oct 30.
- FTC (Federal Trade Commission). 2024. FTC announces crackdown on deceptive AI claims and schemes, Sept 25.
- Franzen C. 2023. Hollywood actors' strike ends with deal to 'protect members from the threat of AI.' *VentureBeat*, Nov 8.
- Ge Y, Knittel CR, MacKenzie D, Zoepf S. 2020. Racial and gender discrimination in transportation network companies. *Journal of Public Economics* 190:104205.
- Grynbaum MM, Mac R. 2023. The Times sues OpenAI and Microsoft over A.I. use of copyrighted work. *The New York Times*, Dec 27.
- Hoskins P. 2022. Uber sued in US over sexual assault claims. *BBC News*, July 14.
- Hussain S, Bhuiyan J, Menezes R. 2020. How Uber and Lyft persuaded California to vote their way. *Los Angeles Times*, Nov 13.
- Lander E, Nelson A. 2021. Americans need a Bill of Rights for an AI-powered world. *WIRED*, Oct 8. <https://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/>.
- Malin BJ, Chandler C. 2017. Free to work anxiously: Splintering precarity among drivers for Uber and Lyft. *Communication, Culture and Critique*, 10(2):382–400.
- National Institute of Standards and Technology. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. Online at www.nist.gov/itl/ai-risk-management-framework.
- OSTP (Office of Science and Technology Policy). 2022. Blueprint for an AI bill of rights: Making automated systems work for the American people. The White House.
- Pasquale F. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Harvard University Press.
- Pollman E, Barry JM. 2017. Regulatory entrepreneurship. *Southern California Law Review* 90:383–448.
- Price M. 2015. Bill before Utah governor to regulate ride-hailing companies. *Associated Press*, March 26.

- Rahman KS, Thelen K. 2019. The rise of the platform business model and the transformation of twenty-first-century capitalism. *Politics & Society* 47(2):177–204.
- Riordan M, Hoddeson L. 1998. *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age*. W. W. Norton & Company.
- Selbst AD, Barocas S. 2018. The intuitive appeal of explainable machines. *Fordham Law Review*, 87(3):1085–139.
- Shaheen S, Cohen A, Zohdy I. 2016. Shared mobility: Current practices and guiding principles (Report No. FHWA-HOP-16-022). U.S. Department of Transportation, Federal Highway Administration.
- Shapiro C, Varian HR. 1998. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.
- Terranova T. 2000. Free labor: Producing culture for the digital economy. *Social Text* 18(2):33–58.
- Thelen K. 2018. Regulating Uber: The politics of the platform economy in Europe and the United States. *Perspectives on Politics* 16(4):938–53.
- USPTO (US Patent and Trademark Office). 2024. USPTO issues guidance concerning the use of AI tools by parties and practitioners, April 10. Online at www.uspto.gov/about-us/news-updates/uspto-issues-guidance-concerning-use-ai-tools-parties-and-practitioners.
- Wiessner D. 2024. Workday must face novel bias lawsuit over AI screening software. Reuters, July 16.
- Winner L. 2021. The democratic shaping of technology: Its rise, fall and possible rebirth. *Engaging Science, Technology, and Society* 7(1). Online at <https://doi.org/10.17351/ests2021.825>.
- Zuboff S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

AI systems' risk preferences should be consistent with those of humans, particularly in critical domains like healthcare and national security.

Alignment of AI Systems' Risk Attitudes, and Four Real-Life Examples

Elisabeth Paté-Cornell



Elisabeth Paté-Cornell (NAE) is a professor of engineering at Stanford University.

Artificial intelligence (AI) systems perform two kinds of functions: gathering and processing information, and, in some cases, suggesting or directly implementing decisions based on that information. With respect to preferences, information should be neutral; decisions are not. The focus in this article is on decisions. The question of AI alignment is how to design and use AI to make or support risk management decisions under uncertainty so that, in the end, the system's recommendations fit the preferences of the decision-maker if they did not in the first place (Paté-Cornell 2024).

AI Preferences and Decision-Making

Preferences have several dimensions, including trade-offs among decision outcome attributes¹ as well as a risk attitude. Imagine, for instance, a medical situation in which an AI system recommends a test that the patient may not want, or a military attack in which drones are guided by an AI system without human intervention.² The AI system may have different approaches to its decisions under uncertainty, but they reflect a risk attitude one way or the other. That risk attitude has to match that of the decision-maker.

It is assumed here that the AI algorithm provides rational decisions based on a decision analysis framework and the von Neumann axioms, often supported by a utility function embedded by the decision analysts within the algorithm (Russel and Norvig 2021). Furthermore, it is assumed that the outcomes x of all scenarios are described by a single unit, like a monetary currency, and a single utility function. In the embedded utility function $U(x)$, the risk attitude at any given level of potential loss x is the negative of the ratio of the second to the first derivative—i.e., $-U''(x)/U'(x)$. In that framework, the optimal option for the decision-maker is that which maximizes his or her expected utility (Abbas and Howard 2016). If the outcomes involve several attributes, the risk attitude may not be the same for all of them. One simple solution adopted in this article is to measure all attributes with the same unit (e.g., US dollars) to obtain a single input in the considered utility function.

Biases can occur in the selection of training data and the processing of information.

Note that the facts involved in the AI-generated information are assumed to be free of preferences, even though, in reality, many types of biases may affect both the information as gathered and processed and the decisions that it supports. Biases can occur, for instance, in the selection of the training data, their processing by the algorithm, and their interpretation by the decision-maker. The information in the AI system could thus be right or wrong, but the facts—if not the judgment of the facts—are generally verifiable. Yet, it is clear that information biases may

¹ Attributes of the decision outcomes could be, for instance, environmental quality and financial cost. The risk attitude for each attribute does not have to be the same. For example, some people may be more risk averse for environmental quality than money, and others may feel the reverse.

² Note that the US Department of Defense does not allow this kind of operation of its forces without human control.

Biases can occur, for instance, in the selection of the training data, their processing by the algorithm, and their interpretation by the decision-maker.

distort the decision process, and step one is to check that the information is right.

In that respect, the case of decisions is different because there is no “right” risk attitude, and each decision-maker may feel differently. That is why the relevant risk attitude has to be determined as a function of specific decisions and of who makes them (Hassan et al. 2024).

Differences in risk attitudes between the human decision-makers and the AI system that gives them recommendations prevent relevant AI advice.

The AI system, however, may not involve the relevant risk attitude for several reasons. The analyst may not know who is or will be the decision-maker, the system may be defined to advise a group of people who may not have the same utility, or the AI prescription may reflect crowd decisions that have been made in the past in a different environment.³ In that case, the problem is to align the AI risk attitude with that of the anticipated decision-maker(s) or affected organization. This requires encoding a common risk attitude, then, if appropriate, accessing the risk attitude factor of the AI system and aligning it to fit the decision-makers’.

Addressing the alignment problem thus requires at least two things: accessing the risk attitude encoded in the AI system (Russel and Norvig 2021) and knowing that of the decision-maker(s), with the understanding that preferences may vary over time. For a decision-maker to adopt the decisions of the AI algorithm, these decisions need to be guided by his or her risk attitude. One

³ Several decision-makers may be involved in different group structures, which will determine their risk attitude. In a classical organization, it is the leadership team that will decide which risk it wants to take. In a less structured but still organized group, decisions may emerge from a vote. In an amorphous group, the risk attitude may depend on the source of information (e.g., social networks, the political climate, and general feelings of the crowd).

Differences in risk attitudes between the human decision-makers and the AI system that gives them recommendations prevent relevant AI advice.

may thus have to adjust that of the AI algorithm. To do so, one needs to have access to that factor as encoded by the decision analyst and the ability to modify it.

The problem of AI systems alignment for different users has been considered in the last few years from different perspectives (Caburao 2025; Gent 2023; Iason 2020; Ransbotham et al. 2022; Wixom et al. 2020). In particular, the alignment problem has been described as being at the interface of risk analysis and decision analysis (Abbas and Howard 2016). Maarten Sap, for instance, tests the social intelligence and the capacity for AI systems to interact with society at large (Sap 2024). This article focuses mostly on individual actors, although it encompasses the interaction between AI recommendations and society, considered as a mega decision-maker.

The alignment of the AI system is relevant to a variety of real-life situations.

One way to describe the alignment problem and its importance in real life is to consider actual situations in which decisions under uncertainty are guided by AI systems. In what follows, four types of situations are described that involve both human and engineering elements: 1) A patient's decision of whether or not to take a medical test; 2) A national security situation involving attack drones; 3) Sports decisions with a specific application to sailing races; and 4) The design of autonomous vehicles, considering their response to different conditions and actors on the road. In each of these cases, what matters most are the circumstances of the decision, the options, the uncertainties, and the risks involved, including their possible consequences. In each case, a decision analysis can reveal the effects of the risk attitudes and of potential discrepancies, which justify the alignment of the AI system.

Case #1: A Medical Test

The AI system can assess the information but not make the decision to take the test given the risks.

Assume that you are a witness in a doctor's office.⁴ A patient had agreed to a mammography, and the result was essentially negative, except for a minor problem with the imaging technology. She agreed to a needle biopsy, which came back negative, but there was still a little fuzziness in the image. An AI system was programmed to provide a recommendation (AlSamhori et al. 2024), which, in this case, was to proceed to further testing via a surgical test. That AI advice was based on the image and its mild uncertainty, as well as the age of the patient. The system was risk averse to protect patients and apparently made its decision based on a risk to the general population. The patient realized that she was less risk averse than the AI system and that she had additional information about herself. She declined further testing based on her assessment of her own health condition and on a second medical opinion. Twenty-five years later, she is still free of cancer, and her decision turned out to be the right one.

In this case, the patient's decision relied on several factors. She was convinced that although the AI system generally knows better about the risks and the benefits involved, she had a clearer knowledge of her own case: her health, her medical history, and the medical history of her family—and thus of the chances *a priori* that she had breast cancer. Also, she was sensitive to the risks of an invasive surgical test, the possibility of infection, and the pains of recovery.

Note also that some AI recommendations regarding cancer treatments, such as those of the IBM Watson system, have been wrong in the past (Lohr 2021). The patient was not in the best position to assess the chances of a false negative, but she had the benefit of the opinion of an expert who could access the image. Based on that information, the patient was willing to face later the remote possibility of a cancer that could have been avoided by early detection. These uncertainties, the patient's valuation of the various outcomes and of the trade-offs involved, her time value (discounting), and her risk attitude were thus at the core of her decision.

The patient needs to know how the AI recommendation was made.

Although the patient had the benefit of an expert opinion, to seriously consider the AI system's surgery rec-

⁴ This is an adaptation of a real case.

ommendation, the patient needed to know the basis of the AI-generated information related to mammography (Branco et al. 2024), as well as the preferences that guided the recommendation (Fryer 2024). But that information is seldom available in the medical field, where AI and large language models are increasingly used, both for diagnosis and prescription (Khalifa et al. 2024; Magrabi et al. 2019; Rajpurkar et al. 2022). Most people may not understand the difference between the facts that guide the AI reasoning and the (arbitrary) preferences encoded in the prescription system. Therefore, patients may simply adopt whatever the algorithm recommends, based on the fact that it generally knows more than they do, but they seldom consider the possibility that the system may have a risk attitude different from theirs. Increasing patients' access to the model parameters and patient education would result in better alignment of the AI system and strengthen the system's role in patients' decision-making.

Case #2: The Use of Autonomous Drones in Combat

The United States does not allow the use of autonomous drones in combat, but some of its adversaries do.

Most drones are used in the US armed forces for reconnaissance and for attacking targets under human judgment. There is thus a human in the loop, both in the attack and in the response that may follow. By contrast, autonomous drones and other lethal autonomous weapons systems are programmed to make decisions without direct human involvement when anticipating threats, identifying targets, and attacking enemy positions (Garamone 2023; Springer 2013). Yet, the distinction between controlled and autonomous drones is, in fact, somewhat ambiguous.

The US Department of Defense (DoD), although it has developed autonomous drones, requires human judgment in combat. China has also developed unmanned combat aerial vehicles that can perform both reconnaissance missions and precision strikes. Such systems could make attack decisions, which could become critical in Chinese operations in the South China Sea. By contrast, other countries, such as Turkey, Russia, and Iran, have actually used fully autonomous drones in combat.

Should the United States change its rules of operations?

From the US perspective, the question for the moment is how to deal with a classic attack situation using controlled drones as opposed to fully autonomous ones, when some US enemies do use fully autonomous drones, which would allow them to react instantaneously to the

detriment of the United States. In the future, the United States may thus face a critical decision about using fully autonomous forces to match the attack of an enemy using such drones.

When using autonomous drones, some adversaries have a critical advantage. Should the United States do so?

Two critical questions will thus have to be addressed by the United States regarding AI-guided military drones: 1) How should one use conventional drones (e.g., when should one delay a counterattack)?, and 2) Should there be a human in the loop, or should the United States allow the use of fully autonomous drones?

The advantages of autonomous drones in combat may be clear, but they will have to be balanced against their potential risks, especially in front of an adversary that is also using them. The autonomous drone might simply be programmed to destroy everything it can, and given the uncertainties, it could destroy an innocent target or, on the contrary, it could lose a target for being too cautious. And like all other autonomous systems, it might simply make a mistake and, for instance, attack its own forces.

Consider, first, the current possibility of an enemy using autonomous drones when the US military is not allowed to do so. After an attack, a US commander has to decide how and when to respond in a situation where attacks can be extremely fast, and speed is critical. That commander may want to launch an immediate counter-attack since the conventional forces may be devastated otherwise. Assume that the commander is at a level of authority where it is necessary to consider the national effects of his or her decision, in addition to the immediate tactical ones. The commander may want to take some time to make a strategic decision, considering the long-term consequences of an immediate reaction to try to decrease the risk of escalation. Since an unchallenged enemy attack can be devastating to the United States, the commander may thus face a trade-off between a minor delay that could allow diplomacy to lead to a truce and an immediate response. In that situation, a key decision is thus that of the commander to respond right away with the available forces or to voluntarily delay that response to allow for diplomatic negotiation or another military option. This assumes that the enemy attack does not reflect a software problem in their own drones, which the commander may not be able to suspect.

In the future, the DoD may thus have to make a fundamental strategic decision of whether or not to use fully autonomous drones and let an AI system guide immedi-

ate automatic responses to an attack. On the one hand, by doing so, the US military might match—or surpass—the enemy forces. On the other hand, it may lose the opportunity of a strategic delay for rational thought in pursuit of combat.

The information that will guide these decisions is obviously not available to the public, or even perhaps to the DoD at this time, but the fundamental formulation of the problem is clear. Under the current circumstances, a commander may or may not want to follow or even access an AI system that could yield useful advice but might not reflect his or her risk attitude and preferences in combat operations. If an AI system is to guide attacks by autonomous drones without direct human intervention, it must be devised so that these attacks can be observed and stopped by a decision-maker according to his or her preferences and risk attitude.

What is the effect of an AI-controlled drone system on deterrence?

Whether fully autonomous drones enhance or decrease attack deterrence is not obvious. On the one hand, they may give incentives to the adversary to slow down, considering the losses that it may incur from an immediate, well-targeted US response. On the other hand, autonomous drones may trigger attacks that could be avoided or delayed for a better chance of an expeditious move towards a truce. There is also the possibility of the enemy hacking the United States' AI and rendering it useless, but it is difficult to "hack" humans, making classic systems more reliable in that respect. In any case, the commanders must be able to stop automatic AI combat systems, which have to be aligned to their preferences to permit choices that may limit destruction on both sides. But one must recognize that this ability may render the United States' AI system vulnerable and ineffective.

Case #3: AI in Sailing Races

There is recent evidence of the value of superior AI systems in regattas.

The victory of the Emirates Team New Zealand over its adversaries in the America's Cup of October 2024 (EmiratesTeamNZ 2024; Goodhart 2021) is a recent demonstration of the value of a superior AI system in a major regatta. There were actually two distinct applications of AI, one in the design of the boat and the other as an operational advisor to the skipper. The boat was an AC75 foiling monohull yacht that included a futuristic hydrofoil designed with the help of AI, which allowed the

boat to fly above the water and travel at over 100 km/h (Gladwell 2021). The AI system on board was autonomous and adapted to the boat but the skipper still had some critical decisions to make. The skipper gets information from that system but often still faces uncertainties about possible variations of the actual weather, as well as the position, speed, and strategy of the competitors. Therefore, he or she may have to make decisions under uncertainties that vary constantly, given his or her main objective about what rank to aim for in the race and what risk to take.

Should the skipper follow the advice of the AI system or account for his or her immediate instincts?

The skipper can consider two options: to trust the AI system and follow its recommendations to shift or not shift the course of the boat, or to tighten or relax the sails (Alves et al. 2017; Anderson et al. 2023; Tagliaferri et al. 2016; Tagliaferri and Viola 2017). The alternative is to keep the boat on its course as planned. In this decision, the skipper faces several sources of uncertainty: the accuracy of the AI sensors that provide information and guidance but may be failing or still learning; changes in the environmental conditions, such as wind and waves, that the AI system may not have anticipated; the moves and the strategy of the competitors; and, finally, the risk of collision with other boats in the race, for instance, around a buoy, which the skipper needs to manage carefully when choosing a course and setting the sails.

The skipper may have the option to specify his or her risk attitude at the time of the design of the AI system or before using it, but that risk attitude may change during the race—for instance, with unexpected performances of the competitors. Yet, a good alignment based on experience can be set at the onset of the race. In the case of New Zealand's victory in the 2024 America's Cup, the owner, the skipper, and, to some degree, the crew were involved from the beginning in the design of the AI system, including its role in the design of the boat and the choice of maneuvering options. All uncertainties were not resolved, but there was a good alignment of preferences in a top-class AI system.

Will AI kill the legitimate competition of sailing races?

All boats, however, may not be guided by an AI algorithm of that power, even if they have AI support (Scuttlebutt Sailing News 2023; van Aartrijk 2002). One may wonder at what point the sport aspect of regattas will be dominated by technology rather than the skills of the crew.

Meanwhile, in addition to AI guidance, the boats will still have to be able to follow the instincts and the skills of the skipper, given good but imperfect AI-generated information, the skipper's assessment of the key factors in the race, and his or her risk attitude. The relevance of the AI system, as in previous cases, thus depends on its alignment with the general risk attitude of the skipper. The capabilities of the top existing AI systems are confidential and carefully protected. Therefore, one does not know exactly what dynamic adjustments will be feasible, but improvements of their alignment and possible changes of preferences can give these boats even greater power than sophisticated algorithms with fixed preference parameters.

Case #4: Autonomous Vehicles

The AI system for AVs should be designed to accommodate a variety of human drivers and customers.

Fully autonomous vehicles (AVs)—as opposed to automatic ones—are totally guided by artificial intelligence; hence, there is a risk of AI malfunctions that may affect passengers (Ajenaghughrure 2020; Thomas 2024; Tong et al. 2023; van der Smagt 2021). The AI system makes decisions that would otherwise be made by a human driver when the AV drives among conventional vehicles (CVs). Comparing risk attitudes is thus complex because of the diversity of drivers, their behaviors, and their risk-taking, especially in situations that may lead to a rear-ending accident. When the AI system works as it should, the AVs follow the rules of the road, whereas some of the CV drivers may not. The AV risk management decisions thus have to be anticipated by the AI analyst when setting the rules of driving in a number of situations, including cars driven as expected or at different speeds and other drivers obeying or not obeying the rules of the road. Similar problems may also occur when AVs, obeying all the rules, do not drive as expected by other drivers.

Managing the risk of an AV accident is complex.

The designer of the AI system thus has to manage the risk of accidents involving AVs. First, technical accidents can be caused by a malfunction of the sensors, actuators, and vision capabilities of the AV (Wang 2022). For example, in 2023, an AV nearly drove into a trench at a construction site in San Francisco (The San Francisco Standard 2023). More generally, AVs can be challenged by poor visibility and other vehicles. These recognition issues are engineering problems that can be addressed as such. But most accidents of AVs in mixed traffic are rear-ends

caused by CVs (Cunneen 2023). Again, this may have to do with differences in speed, especially in areas where the speed is limited. The risk attitude of the AV can perhaps be automatically modified if it perceives an immediate danger, but that flexibility may create another risk if the system malfunctions. In many cases, educating common drivers so that they are familiar with the functioning of AVs could help reduce risk (Dragomir et al. 2024).

Similarly, the passenger who decides to ride in an AV is the one who takes the risk inherent to the vehicle, and many people are still uncertain about taking that risk (Naiseh et al. 2024). One can perhaps imagine a system in which the passenger has access to the guidance system through a knob on the board that would allow him or her to post a risk aversion level and to align it to that of the AI system. However, there might be liability issues if the AV got into an accident when the knob was set at a high-risk tolerance.

Conclusion

When using an AI system in risk management, the challenge is to ensure its alignment with the risk attitude of the user whenever feasible.

The decision-maker, supported by an AI system, still plays a major role in the choice of management options. A patient advised by an AI system should know the source of the information to be able to make a personal decision. A commander who operates automatic combat drones has to be able to override the AI decisions based on the knowledge he or she gained from wargaming. The skipper who runs in a race can be trained on a simulator and acquire the skills that allow him or her to adapt the AI system's decisions to the circumstances. And the designer of an autonomous vehicle should account for the risk attitude of the expected customers when designing the safety system.

***When using an AI system
in risk management, the
challenge is to ensure its
alignment with the risk attitude
of the user whenever feasible.***

One of the analytical challenges is to consider the dynamics of the situation and the variations of risk attitudes over time. To do so requires that the analyst know the decision-maker's preferences when designing the algorithm. If the situation requires it, that alignment must be adjusted at decision time, which will require transparency and flexibility of the AI system. This will imply allowing for disclosure and possibly adjustment of its risk attitude for a specific decision-maker or a homogeneous group of people affected.

Acknowledgment

The author is most grateful for the advice of Professor Peter Glynn and of Admiral James O. Ellis, both members of the NAE.

References

- Abbas A, Howard R. 2016. *Foundations of Decision Analysis*. Pearson.
- Ajenaghughrure IB, da Costa Sousa SC, Lamas D. 2020. Risk and trust in artificial intelligence technologies: A case study of autonomous vehicles. In: 13th International Conference on Human System Interaction, 118–123. IEEE Computer Society.
- AlSamhori JF, AlSamhori ARF, Duncan LA, Qalajo A, Alshahwan HF, Alabbadi M, Al Soudi M, Zakraoui R, AlSamhori AF, Alryalat SA, and 1 other. 2024. Artificial intelligence for breast cancer: Implications for diagnosis and management. *Journal of Medicine, Surgery and Public Health* 3:1–10.
- Alves B, Veloso B, Malheiro B. 2017. An Agent-Based Platform for Autonomous Sailing Research and Competition. In: *Robotic Sailing 2017: Proceedings of the 10th International Robotic Sailing Conference*, 31–38. Øvergård KI, ed. Springer.
- Anderson J, Sithungu S, Ehlers E. 2023. Route optimization for sailing vessels using artificial intelligence techniques. In: *Proceedings of the International Conference of Computational Intelligence and Intelligent Systems*, 60–66. Association for Computing Machinery.
- Branco PESC, Franco AHS, de Oliveira AP, Carneiro IMC, de Carvalho LMC, de Souza JIN, Leandro DR, Cândido EB. 2024. Artificial intelligence in mammography: A systematic review of the external validation. *Rev Bras Ginecol Obstet*. Online at doi 10.61622/rbgo/2024rbgo71.
- Caburao E. 2025. Leveraging AI in risk management for effective implementation. *Safety Culture*, Feb 12.
- Cunneen M. 2023. Autonomous vehicles, artificial intelligence, risk and colliding narratives. In: *Connected and Automated Vehicles: Integrating Engineering and Ethics*. Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 67, 175–95. Fossa F, Cheli F, eds. Springer.
- Dragomir D, Popișter F, Kabak KE. 2024. Using AI tools to enhance the risk management process in the automotive industry. In: *Advances in Manufacturing IV*, volume 2 (Lecture Notes in Mechanical Engineering), 189–98. Trojanowska J, Kujawińska A, Pavlenko I, Husar J, eds. Springer.
- EmiratesTeamNZ. 2024. The Emirates Team New Zealand Story. Youtube, Oct.
- Fryr N, Groß D, Lipprandt M. The ethical requirement of explainability for AI-DSS in healthcare: A systematic review of reasons. *BMC Medical Ethics* 25: article 104.
- Garamone J. 2023. DOD Updates Autonomy in Weapons System Directive. DoD News, US Department of Defense, Jan.
- Gent E. 2023. What is the AI alignment problem and how can it be solved? *New Scientist*, May 10.
- Gladwell R. 2021 America's Cup: Emirates team New Zealand use artificial intelligence to find the fastest way. *Sail-World*, March 4.
- Goodhart B. 2021. How team New Zealand did the unexpected with its America's Cup boat. *GQ*, March 10.
- Hassan N, Slight R, Bimpong K, Bates D, Weiand D, Vellinga A, Morgan G, Slight S. 2024. Systematic review to understand users' perspectives on AI-enabled decision aids to inform shared decision making. *npj Digital Medicine* 7: article number 332.
- Iason G. 2020. Artificial intelligence, values and alignment. *Minds and Machines* 30:411–37.
- Khalifa M, Albadauy M, Iqbal U. 2024. Advancing clinical decision support: The role of artificial intelligence across six domains. *Computer Methods and Programs in Biomedicine* 5: article number 100142.
- Magrabi F, Ammenwerth E, McNair JB, Dde Keizer NF, Hypönen H, Nykänen P, Rigby M, Scott PJ, Vehko VT, Wong S-Y, Georgiou A. 2019. Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications. *Yearbook of Medical Informatics* 28(1):128–34.
- Naiseh M, Clark J, Akarsu T, Hanoach Y, Brito M, Wald M, Webster T, Shukla P. 2024. Trust, risk perception, and intention to use autonomous vehicles: An interdisciplinary bibliometric review. *AI & Society*. Online at <https://doi.org/10.1007/s00146-024-01895-2>.
- Paté-Cornell E. 2024. Preferences in AI algorithms: The need for relevant risk attitudes in automated decisions under uncertainties. *Risk Analysis* 44(10):1–7. Online at <https://doi.org/10.1111/risa.14268>.
- Rajpurkar PE, Chen E, Banerjee O, Topol EJ. 2022. AI in health and medicine. *Nature Medicine* 28:31–8.

- Ransbotham S, Kiron D, Candelon F, Khodabandeh S, Chu M. 2022. Achieving Individual- and Organizational-Value with AI, Findings from the 2022 Artificial Intelligence and Business Strategy Global Executive Study and Research Project. MIT Sloan Management Review.
- Russel S, Norvig P. 2021. Artificial Intelligence: A Modern Approach, 4th edition, Pearson Series in Artificial Intelligence. Pearson.
- Sap M. 2024. Artificial social intelligence? On the challenges of socially aware and ethically informed large language models. *The Bridge* 54(4):20–23.
- Springer PJ. 2013. Military Robots and Drones: A Reference Handbook. ABC-CLIO.
- Tagliaferri F, Hayes B, Viola IM, Djokic S. 2016. Wind modeling with nested Markov chains. *Journal of Wind Engineering and Industrial Aerodynamics* 157:118–24.
- Tagliaferri F, Viola IM. 2017. A real-time strategy-decision program for sailing yacht races. *Ocean Engineering* 134:129–39.
- Thomas J. 2024. Integrating machine learning and AI in automotive safety. *International Journal of Innovative Science and Research Technology* 9(1). Online at <https://doi.org/10.5281/zenodo.10670478>.
- Tong K, Guo F, Solmaz S, Steinberger M, Horn M. 2023. Risk monitoring and mitigation for automated vehicles: A model predictive control perspective. In: 2023 IEEE International Automated Vehicle Validation Conference (IAVVC), 1–7. IEEE.
- van Aartrijk ML, Tagliola CP, Adriaans PW. 2002. AI on the ocean: The RoboSail Project. In: Proceedings of the 15th European Conference on Artificial Intelligence, 653–57. IOS Press.
- van der Smagt P. 2021. Artificial intelligence in the automotive industry. Joint Aida-Tran Hearing on AI and Transportation, EU Transport Policies. How to prepare for AI while minimizing risk.
- Zigoris J. 2023. Driverless Waymo self-driving car almost digs itself into hole—literally. *The San Francisco Standard*, Jan 15.
- Wang D, Fu W, Song Q, Zhou J. 2022. Potential risk assessment for safe driving of autonomous vehicles under occluded vision. *Sci Rep* 12: article number 4981.
- Wixom B, Someh I, Gregory R. 2020. AI alignment: A new management paradigm. Research Briefing. MIT Center for Information Systems Research.

Ingestible devices present a novel approach to accessing the gastrointestinal tract, making possible therapeutic interventions that can modulate gut function and potentially improve health outcomes.

Unlocking the Gut's Brain with Ingestible Bioelectronics

Shriya Srinivasan



Shriya Srinivasan is an assistant professor of bio-engineering at the Harvard School of Engineering and Applied Sciences and the director of the Harvard BIONICs lab, where she develops neuro-technology.

With the increasingly popular use of weight loss drugs, such as Ozempic and Wegovy (Sodhi et al. 2023), there has been a surge in the occurrence of gastroparesis, a gastrointestinal disorder wherein the stomach experiences paralysis, creating symptoms of nausea, vomiting, and abdominal pain. The prevalence and debilitating effects of gastroparesis have not only underscored that these drugs may not be the comprehensive solution we seek but have also raised the urgent need for us to rigorously understand such gastrointestinal disorders and develop effective treatments. At its core, this requires unraveling the intricate interactions between the gut and its nervous system—a relationship that is emerging as a significant frontier in medical research.

The enteric nervous system (ENS), often called the gut's brain, is a complex network of 400-600 million neurons embedded in the walls of the GI tract, running from the esophagus to the anal canal. As the largest and most complex part of the peripheral nervous system, it directs digestive processes and interacts with various organs to influence overall well-being. Researchers have uncovered that its complex functions among neural, hormonal, and microbial axes affect everything from mood and cognition to immunity and energy balance (Kaelberer et al. 2018). However, because of its highly interconnected nature with other organ systems, isolating its role has been challenging. Further, the intricate anatomy of the tract presents unique challenges in the design and engineering of devices that can interface with the ENS. Specifically, the nerves are embedded in a weblike pattern along the length of the gastrointestinal tract, which undergoes significant motion during peristalsis, the wave-like contraction of muscles lining the digestive tract. Ideally, we would be able to probe each segment of the tract (esophagus, stomach, intestines) and track its neural activity throughout the day, in response to activity, meals, diseases, and other variables, to understand the functioning of the ENS.

Ingestible devices present a novel approach to access the gastrointestinal tract; with a form factor no larger than an everyday multivitamin, they allow us to record and stimulate the ENS at various points. This capability not only opens new research avenues but also paves the way for therapeutic interventions that can modulate gut function and potentially improve health outcomes.

Given the unique properties of the GI tract, the design of an ingestible device requires creativity and a holistic approach. First, it must fit into a pill-sized form factor to be ingested orally and traverse the tract without causing obstruction or perforation. The surface of the device must withstand the pH of each segment and be resistant to the contents of the tract. Additionally, to effectively record and stimulate the ENS, devices must interact with the gut's mucosal or muscular layers, anchoring themselves or having a robust contact mechanism. Finally, power and communication must be sufficiently miniaturized to be carried onboard or overcome the depth within the body to be wirelessly controlled by external devices. Early work on capsule endoscopy systems paved the way for the first ingestible devices, transforming our ability to visualize and diagnose GI tract conditions non-invasively (Iddan et al. 2000). These devices set design benchmarks in size and materials, crucial for safely navigating the harsh environment of the gastrointestinal tract.

What neural interfacing capabilities can be achieved with ingestible devices?

My research in Professor Giovanni Traverso's lab at MIT delved into the possibilities from various angles to explore the challenges and potential of multimodal ingestible systems, specifically designed to interrogate the ENS.

Chemical Actuation: Drugs, including serotonin and dopamine receptor agonists/antagonists, can directly modulate the ENS, which affects motility and the sensory functions, leading to nausea and vomiting. Evolving from dissolvable tablets and temperature-triggered materials to self-deploying needle systems for administering macromolecules directly into the gastrointestinal lining, the drug delivery field has seen dramatic advancements over the past few decades (Abramson et al. 2019). Incorporating mechanical and chemical elements, I recently developed the RoboCap. This ingestible utilizes an onboard motor to rotate miniature fins that clear mucus, similar to a tunnel-digging device, exposing the intestinal surface to allow the enhanced absorption of drugs, especially of large molecules (Srinivasan et al. 2022). These types of functionalities bolster the capability to deliver drugs in a targeted fashion and with controllable release kinetics.

Electrical Actuation: ENS activation via stimulation has been previously achieved via unique geometries that prioritized high surface area contact with the lumen, the inner surface of the tract, or microneedle penetration into the muscle (Abramson et al. 2020; Ramadi et al. 2023; Srinivasan and Dosso et al. 2024). In conditions like ileus, intestinal paralysis, the entire tract must be stimulated. To achieve stimulation along the length of the intestines, my team and I designed the ingestible self-propelling device for intestinal reanimation (INSPIRE). This device opens up into an S-shaped device after reaching the intestines and performs electrical stimulation through four contacts at the edges, supported by an onboard power source and microcontroller. As the intestines contract in response to stimulation, the device undergoes a shape change and springs forward to stimulate the adjacent segment. By incorporating mechanical and electrical stimulation, this device improved intestinal motility by up to 140% and decreased mean passage time from 8.6 days in controls to 2.5 days. In addition to neuromuscular activation for motility, stimulating the ENS may allow us to modulate the neural pathways to trigger hormonal release.

Mechanical Actuation: Another way to interact with the ENS is to trigger its mechanoreceptors with mechanical stimulation. In the case of obesity, I saw an interesting opportunity to tap into the hunger-satiety reflex circuit to harness the body's own mechanisms to solve the issue of overeating. The vibratory ingestible bioelectronic stimulator (VIBES) is an ingestible device that vibrates the gastric musculature, artificially activating distension receptors. These artificially signal to the brain that the stomach is full—creating an illusory satiety. Swine treated with this pill felt early satiety and significantly decreased food intake (by roughly 40%). Interestingly, this also caused the release of hormones consistent with feeding (Srinivasan et al. 2023). By tapping into such neural reflex circuits, this device offers a mechanistic solution for obesity, which is also cost-effective, greatly increasing access to therapy.

Sensing: The ability to directly record from the ENS will be critical in improving our understanding and diagnosis of neural pathologies (Srinivasan and Liu et al. 2024). The multimodal electrophysiology via ingestible, gastric, untethered tracking (MiGUT) device rolls out a sensing electrode ribbon to make contact with the mucosa (You et al. 2024). The device then records and wirelessly transmits biopotentials to an external receiver. Similar embodiments extending these concepts to all segments of the GI tract will empower the development of analytical pipelines and electrophysiologic biomarkers for neurally mediated GI pathologies. Apart from electrical sensing, further research is required to develop robust mechanical, chemical, and molecular sensing strategies and to integrate these to provide a continuous readout of GI function along the tract. This would enable correlations between neural signals and disease states.

Conclusion

Wearable devices have fundamentally changed how we live. We can record our heart activity to curate workouts, granularly track our sleep to optimize restfulness, and even detect falls or warn us of noise levels. Similarly, I envision a future in which ingestible devices will be able to monitor the tract and responsively stimulate it to allow for advanced insights and treatments for gastrointestinal health. For example, in the context of gastroparesis, an ingestible device might reside in the stomach or intestines, continuously monitoring the pH of the contents along with the electrogastrogram (EGG), which provides

I envision a future in which ingestible devices will be able to monitor the tract and responsively stimulate it to allow for advanced insights and treatments for gastrointestinal health.

electrical signals from the luminal surface. When the pH increases, as food enters the tract, if the EGG reflects depressed motility, then the device could electrically stimulate the tract to excite the ENS locally and promote peristalsis. Alternatively, a network of devices could also implement a responsive system and be able to coordinate function in the independent but linked segments of the tract. For example, devices monitoring the electrical activity of the intestines would survey for abnormal spasms and then deliver drugs at proximal and distal sites to reduce the unwanted muscular activity.

We are still in the early days of development for such ingestible devices, but the future holds promise. The gut has been proposed as a crucial early biomarker for various diseases. As research in the microbiome and neurogastroenterology continues to expand, the potential of leveraging data from ingestible bioelectronics becomes increasingly compelling. These devices, designed to traverse the gastrointestinal tract, must meet rigorous engineering standards to withstand the challenging environment, maintain contact for effective monitoring and stimulation, and handle power and data transmission demands. Further, it is critical to address challenges relating to device disposal and environmental impact to ensure that sustainability considerations are sufficiently integrated into the development process. Ultimately, the future of ingestible bioelectronics will lie not only in their technological innovation but also in their potential to seamlessly integrate into and enhance existing medical practices and lifestyles. By advancing precision health monitoring and creating tailored therapeutic interventions, these devices offer a promising pathway toward significantly improving patients' quality of life.

References

- Abramson A, Caffarel-Salvador E, Khang M, Dellal D, Silverstein D, Gao Y, Frederiksen MR, Vegge A, Hubálek F, Water JJ, and 13 others. 2019. An ingestible self-orienting system for oral delivery of macromolecules. *Science* 363(6427):611–15.
- Abramson A, Dellal D, Kong YL, Zhou J, Gao Y, Collins J, Tamang S, Wainer J, McManus R, Hayward A, and 6 others. 2020. Ingestible transiently anchoring electronics for microstimulation and conductive signaling. *Science Advances* 6(35):eaaz0127.
- Iddan G, Meron G, Glukhovskiy A, Swain P. 2000. Wireless capsule endoscopy. *Nature* 405(6785):417.
- Kaelberer MM, Buchanan KL, Klein ME, Barth BB, Montoya MM, Shen X, Bohórquez DV. 2018. A gut-brain neural circuit for nutrient sensory transduction. *Science* 361(6408):eaat5236.
- Ramadi KB, McRae JC, Selsing G, Su A, Fernandes R, Hickling M, Rios B, Babae S, Min S, Gwynne D, and 8 others. 2023. Bioinspired, ingestible electroceutical capsules for hunger-regulating hormone modulation. *Science Robotics* 8(77):eade9676.
- Sodhi M, Rezaeianzadeh R, Kezouh A, Etminan M. 2023. Risk of gastrointestinal adverse events associated with glucagon-like peptide-1 receptor agonists for weight loss. *JAMA* 330(18):1795–97.
- Srinivasan SS, Alshareef A, Hwang A, Byrne C, Kuosmanen J, Ishida K, Jenkins J, Liu S, Gierlach A, Madani WAM, and 3 others. 2023. A vibrating ingestible bioelectronic stimulator modulates gastric stretch receptors for illusory satiety. *Science Advances* 9(51):eadj3003.
- Srinivasan SS, Alshareef A, Hwang AV, Kang Z, Kuosmanen J, Ishida K, Jenkins J, Liu S, Madani WAM, Lennerz J, and 5 others. 2022. RoboCap: Robotic mucus-clearing capsule for enhanced drug delivery in the gastrointestinal tract. *Science Robotics* 7(70):eabp9066.
- Srinivasan SS, Dosso J, Huang H-W, Selsing G, Alshareef A, Kuosmanen J, Ishida K, Jenkins J, Madani WAM, Hayward A, and 1 other. 2024. An ingestible self-propelling device for intestinal reanimation. *Science Robotics* 9(87):eadh8170.
- Srinivasan SS, Liu S, Hotta R, Bhav S, Alshareef A, Ying B, Selsing G, Kuosmanen J, Ishida K, Jenkins J, and 5 others. 2024b. Luminal electrophysiological neuroprofiling system for gastrointestinal neuromuscular diseases. *Device* 2(7):100400.
- You SS, Gierlach A, Schmidt P, Selsing G, Moon I, Ishida K, Jenkins J, Madani WAM, Yang S-Y, Huang H-W, and 4 others. 2024. An ingestible device for gastric electrophysiology. *Nature Electronics* 7:497–508.

Op-Ed

Why Engineers Should Learn Economics

Debasis Mitra



Debasis Mitra (NAE) is a senior research scientist in the Department of Electrical Engineering at Columbia University.

Economics and economists wield great influence in policy-making at various levels of both the public and private sectors of our society. Examples in the public sector include policies for industry, trade, and regulations. In the private sector, top managers of engineering firms often find it to their advantage to talk up their economics know-how and economic reasoning in justifying their decisions, even when they lack depth in their knowledge. While economists hold sway in policy-making, engineers are in the unenviable position of being the policy-enablers without a major say in policy-making. Engineers are also typically directly affected by these policies, arguably more so than workers in other professions. Economists and engineers generally have an inadequate understanding of the other profession's basic knowledge and goals, and they have much to gain from changing the status quo.

Economics is understandably essential for policy-setting, and a reason that engineers are largely ignored is their ignorance of the field. It is true that “engineering

Engineers and economists have much to learn from each other, and there is much to be gained from them working collaboratively.

economics” has a long history of support for inclusion in the engineering curriculum (Hayford 1917), but what has been proposed is largely accounting in substance. What is needed is a broad understanding of micro- and macroeconomics, including, at a bare minimum, models for and analysis of social welfare, consumption, investments, markets, growth models, business cycles, and fiscal policy. Decision-makers in engineering have much to gain from such knowledge.

Do economists understand the workings of engineering systems? Take, for example, outsourcing, with globalization at the extreme. Globalization affected society broadly and engineers especially. Through the '80s and '90s, economists were the cheerleaders, making the model-based case for mutual benefits to the outsourcer and the contract manufacturer. Economists now have a better handle on all the implications (Autor et al. 2016; Krugman 2019). Had engineers been at the table when the outsourcing dogma was being built up, then, just possibly, they may have pointed out that their professional experience spoke strongly to the downside, which would have been a counterbalance to the arguments for globalization.

Engineering systems' success depends on tight feedback loops between different working groups, each with its particular specialty. This gives engineers an ingrained and intuitive understanding of the value of such couplings. The Bell System gives us a historical perspective on how these feedback loops worked. Consider Hendrik Bode's career at Bell Telephone Laboratories, during which he worked on mathematics, circuit theory, control systems, military systems, and, after retirement, as a professor of systems engineering at Harvard (Brooks 1989). Bode's 1971 monograph, *Synergy: Technical Integration and Technological Innovation in the Bell System*, champions “integration” of systems. Here is a quote from the book that so well describes the workings of integration: “In writing about the transistor discovery, Dr. M.J.Kelly, then President of BTL, said, ‘In accord with our policy of concentrating the efforts of our scientists on research, we immediately

Economists and engineers generally have an inadequate understanding of the other profession's basic knowledge and goals, and they have much to gain from changing the status quo.

formed a closely associated fundamental development group to acquire that body of technological knowledge essential to the development and design of transistors for the many specific communications applications that would certainly follow. They have interested themselves in such problems as the factors controlling the bandwidth of amplification; the noise figure; the amount of amplification possible per stage; energy levels of output; basic materials, processing and structure studies essential for controlled development and design of transistors for specific functions; etc.” Bode’s book makes the case that systems engineering helps in establishing feedback loops by linking specialized compartments and subsequently smoothing the coordination of their activities. Indeed, critical to engineering in the Bell System was the intimate interworking of systems and services with substantial help from a large contingent of systems engineers at the interfaces (Gilliam 2023).

The breakup of AT&T and the spin-off of Lucent Technologies were based in part on the apparent value of separating “systems” and “services.” There were irreparable losses to both sides, which accumulated over the years. A more recent example of the losses that build up over time as a consequence of breaking the feedback loops is that of Boeing divesting various functions that had been integrated, including fuselage construction to Spirit Aero-systems. The costs so greatly outweighed the benefits that apparently a reunion is now in the works (Sindreu 2024).

Does economic theory provide an understanding of the value of integration in engineering? Does it have a handle on the loss from breaking an existing integration? Surveying the economics literature, the answer would appear to be no, at least not to a degree that is adequate

for the high goal of influencing national and corporate policy. If, on the other hand, engineers understood economics and spoke the language, then at least they would be able to make their case, which could, just possibly, be heard.

It must be said that none of this is open and shut. That is, there are instances where engineering feedback loops have been broken with apparent success. For instance, from about 1970 to at least the turn of the century, Intel exemplified integrated chip design and manufacture. The integrated model started breaking up around 2010, to be replaced by the current model wherein the design is done in-house and manufacturing is outsourced to a foundry contractor, such as Taiwan Semiconductor Manufacturing Company (TSMC), the premier chip manufacturer. So, when do integrated operations out-perform outsourcing? In my opinion, this is a large area of engineering economics with many open questions. This is so even without taking national security into account.

It should be noted that there are areas where engineering experience and economic theory converge. Take, for instance, learning-by-doing. The feedback loops connecting manufacturing and design run continuously to improve performance as measured by various metrics. The Nobel Prize-winning economist Kenneth Arrow modeled and analyzed this process in his classic 1962 paper, “The Economic Implications of Learning by Doing.” The following quote from the paper puts it succinctly: “...to produce the N^{th} airframe of a given type, counting from the inception of production, the amount of labor required is proportional to $N^{-1/3}$. This relation has become basic in the production and cost planning of the United States Air Force.”

No less consequential than cost reduction is the gain in knowledge. It should be noted that when work is outsourced the gain in knowledge capital from learning-by-doing accrues to the doer, not to the outsourcer. The position that TSMC occupies today in the semiconductor business must surely be in part due to the benefits of the knowledge capital that it has accumulated. This too should be part of the economics of outsourcing.

Turning to the future, there would appear to be tremendous scope in engineers and economists working collaboratively to tackle the greatest challenge facing humankind today, climate change. This may not have happened in the past, yet the potential exists. Reasons to be hopeful emerge from looking in the rearview mirror at the occurrence of several related and parallel developments in the engineering and economic sciences.

For instance, in recent years economists have been looking hard at the policy consequences of predictions of future global average surface temperatures due to global warming from the greenhouse effect. There is strong evidence that the probability distribution of the temperature is heavy-tailed, and the variance is infinite, which translates to probabilities of extreme future temperatures being substantially higher than would be the case with thin-tailed distributions. When it comes to policy implications, there is a schism, with one camp (see “dismal theorem” [Weitzman 2011]) championing extreme measures to avoid dire consequences, while another camp downplays the aura around fat tails when it comes down to policy implications, and goes on to promote far less drastic policy responses that put global warming on par with other possible calamities (Pyndyck 2011).

This schism in the economics community mirrors one that enveloped the internet traffic engineering community beginning in the 1990s. Network engineers devoted considerable effort to collecting and analyzing data of internet packet traffic, which led to the discovery that the packet distributions have heavy tails. This discovery led a group of engineers to claim that the established processes of network control and design based on exponential and Markovian assumptions were invalid (Beran et al. 1995; Leland et al. 1994). Just as with the economists, counterarguments followed. Soon thereafter, the parameter domain in network design where such heavy-tail behavior mattered was substantially narrowed in the case of video traffic, the killer application on the internet at the time (Heyman and Lakshman 1966). A post-2000 study by statisticians examining network traffic (Cao et al. 2003) showed that in the scenario of higher-speed links and the multiplexing of many traffic streams, possibly having heavy-tail characteristics, the resulting traffic tends toward independence and exponentiality.

The rationale for describing these debates and schisms in the economic and engineering communities is to take note that, in spite of the considerable differences in context, there is nonetheless commonality in the mindset, techniques, and approaches to problems.

Yet another pair of parallel discussions involves the overhang of history in policy-making, which in economics is captured in the long-established concept of “path dependence.” It has taken on new importance with climate change, since it focuses attention on both the difficulties of weaning ourselves off of past dependencies on fossil fuel and the future consequences of the large investments currently being made in energy generation and storage (Aghion et al. 2014). In engineering systems

there is a similar focus on the role of the past, which manifests when multiple equilibrium states exist and initial conditions determine trajectories to steady state at one of the equilibrium states. These concepts are long-established in engineering, especially in the analyses of nonlinear dynamical systems and control and electrical circuits. In certain cases, multiple equilibrium states exist and control strategies are devised for the system to equilibrate to a preferred state, and in other cases, systems are designed to ensure that a unique equilibrium state exists. This is yet another illustration of affinity and the potential gain in combining the knowledge stocks in the disciplines to tackle problems.

There are other fundamental disciplinary skills that can be brought to bear in tackling climate change. For instance, in problem-solving, both in engineering and economics, an essential step is to settle on the operational time scale, and, unsurprisingly, time scale separation is in both disciplines’ toolboxes. These deep-rooted skills should provide common ground in jointly addressing the grand challenge.

Engineers stand to benefit from being educated in economics. First, they will be better positioned to manage their own businesses. Globalization as we have known it may be diminishing, but the scale of business and, importantly, the need to address grand challenges, such as climate change, will undoubtedly become more global and complex, and knowledge of economic fundamentals will be ever-increasingly essential. Second, it will allow engineers to join economists in debating and setting policy at the higher levels of our society. Also, the skills and tools that are common to both knowledge bases should ease the path for engineers learning economics.

(Note: A fair question is “What have you done to edu-

Engineers stand to benefit from being educated in economics.

cate engineers in economics?” I have created two graduate courses, which I teach at Columbia, that intertwine engineering and economics, “Internet Economics, Engineering and the Implications for Society” and “Future Energy: Economics, Systems, Policies.”)

References

- Aghion P, Hepburn C, Teytelboym A, Zenghelis D. 2014. Chapter 4: Path dependence, innovation and the economics of climate change. In: *Handbook on Green Growth*, 67–83. Fouquet R, ed. Edward Elgar.
- Arrow K. 1962. The economic implications of learning by doing. *Review of Economic Studies* 29:155–73.
- Autor DH, Dorn D, Hanson GH. 2016. The China shock: Learning from labor-market adjustment to large changes in trade. *Annual Review of Economics* 8:205–40.
- Beran J, Sherman R, Taqqu M, Willinger W. 1995. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications* 43(2/3/4):1566–79.
- Bode HW. 1971. *Synergy: Technical Integration and Technological Innovation in the Bell System*. Bell Laboratories.
- Brooks H. 1989. Hendrik Wade Bode. Memorial Tributes. National Academy of Engineering. Online at <https://www.nae.edu/189189/HENDRIK-WADE-BODE-19051982>.
- Cao J, Cleveland WS, Lin D, Sun DX. 2003. Internet traffic tends towards poisson and independent as the load increases. In: *Nonlinear estimation and classification in Lecture Notes in Statistics* (vol. 171), 83–109. Denison DD, Hansen MH, Holmes CC, B. Mallick B, Yu B, eds. Springer.
- Gilliam E. 2023. How did places like Bell Labs know how to ask the right questions? The Good Science Project, April 22. Online at goodscience.substack.com/p/how-did-places-like-bell-labs-know.
- Hayford J. 1917. The relation of engineering to economics. *Journal of Political Economy* 25(1):59–63.
- Heyman DP, Lakshman TV. 1966. What are the implications of long-range dependence for VBR-video traffic engineering. *IEEE/ACM Transactions on Networking* 4(3):301–317.
- Krugman P. 2019. Globalization: What did we miss? In: *Meeting Globalization's Challenges: Policies to Make Trade Work for All*, 113–120. Catao L, Obstfeld M, eds. Princeton University Press.
- Leland W, Taqqu M, Willinger W, Wilson D. 1994. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking* 2:1–15.
- Pyndyck RS. 2010. Fat tails, thin tails, and climate change policy. 2010. National Bureau of Economic Research: working paper 16353. Online at: <http://www.nber.org/papers/w16353>.
- Sindreu J. 2024. Boeing calls time on the great American outsourcing. *The Wall Street Journal*, July 2.
- Weitzman ML. 2011. Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy* 5(2):275–92.

NAE News and Notes

NAE Members Awarded National Medals of Science and of Technology and Innovation

The 2024 National Medal of Science Laureates and 2024 National Medal of Technology and Innovation Laureates were awarded by President Joe Biden on January 3, 2025. Laureates received their medals from **Arati Prabhakar**, director of the Office of Science and Technology Policy, at a ceremony held at the Eisenhower Executive Office Building. The National Medal of Science was established by Congress in 1959 and is administered by the National Science Foundation; the National Medal of Technology and Innovation was established by Congress in 1980 and is administered by the US Patent and Trademark Office.

National Medal of Science Recipients

Angela M. Belcher (NAS), James Mason Crafts Professor of Biologi-

cal Engineering & Materials Science, Massachusetts Institute of Technology: For work designing novel materials for applications that include solar cells, batteries, and medical imaging.

Emery Neal Brown (NAS, NAM), Taplin Professor of Medical Engineering, Massachusetts Institute of Technology: For work that has revealed how anesthesia affects the brain.

Ingrid Daubechies (NAS), James B. Duke Professor of Mathematics, Duke University: For pioneering discoveries in the development of wavelets.

Cynthia Dwork (NAS), Gordon McKay Professor of Computer Science, Harvard University: For visionary contributions to the field of computer science and secure public key cryptography.

National Medal of Technology and Innovation Recipients

Martin Cooper, chairman, Dyna LLC: For inventing the handheld cellular phone and revolutionizing worldwide communications. Martin Cooper delivered breakthroughs for cellular telephone and network technologies that have dramatically altered the world as we know it – changing our sense of proximity to others around the globe, the way we perceive ourselves, and our universe of possibilities.

Eric R. Fossum, John H. Krehbiel Sr. Professor for Emerging Technologies and vice provost, Dartmouth College: For inventing world-changing “camera-on-a-chip” technology that has turned billions of phones into cameras and transformed everyday life. When NASA needed smaller cameras to take into space, Eric Fossum developed a groundbreaking image sensor



Belcher with Prabhakar



Brown with Prabhakar



Cooper with Prabhakar



Daubechies with Prabhakar



Dwork with Prabhakar



Hammond with Prabhakar

and then worked to use it in medicine, business, security, entertainment, and more, while also mentoring legions of young entrepreneurs pushing the bounds of innovation.

Paula T. Hammond (NAS, NAM), Institute Professor, Massachusetts Institute of Technology: For groundbreaking research in nanoscale engineering. Paula Hammond pioneered novel materials that have revolutionized how we deliver cancer drugs to cancer patients and how we store solar energy. An inventor and mentor, Paula has paved the way for a more diverse, inclusive scientific workforce that taps into the full talents of our nation.

Kristina M. Johnson, CEO, Johnson Energy Holdings LLC: For pioneering work that has transformed optoelectronic devices, 3D imaging, and color management systems. Kristina Johnson has channeled her ingenuity and optimism into developing technologies that have improved processes for mammograms and pap smears, promoted clean energy, elevated the entertainment industry, and more – while working to expand the field of STEM for all Americans.

Victor B. Lawrence, Batchelor Chair Professor of Electrical Engineering and associate dean for special programs, Stevens Institute of Technology: For a lifetime of prolific inno-

vation in telecommunications and high-speed internet technology. Victor Lawrence has dedicated his life to expanding the realm of possibilities worldwide. By bringing fiber-optic connectivity to the African continent and improving global internet accessibility, he has enhanced the security, opportunity, and well-being of people around the world.

David R. Walt (NAM), Hansjorg Wyss Professor of Pathology, Harvard Medical School: For setting a new gold standard in genetic analysis that is transforming medical research, care, and well-being. David Walt pioneered the use of microwell arrays to analyze thousands of genes at once



Fossum with Prabhakar



Johnson with Prabhakar



Lawrence with Prabhakar



Walt with Prabhakar

and detect single molecules, making DNA sequencing exponentially more



Yock with Prabhakar

accurate and affordable, and promising simple biomarker blood tests that

may revolutionize our approach to cancer and other conditions – giving people renewed hope.

Paul G. Yock, Martha M. Weiland Professor of Bioengineering/Medicine and director, Program in Biodesign, Stanford University: For innovations in interventional cardiology. Paul Yock's visionary work understanding the human heart is applied around the world today to improve patient care and save countless lives. His creation of the Biodesign approach to training future leaders of biotechnology and health care ensures his insights and experience will benefit generations to come.

Class of 2025 Elected

The National Academy of Engineering (NAE) elected 128 new members and 22 international members to the Class of 2025, announced NAE President **John L. Anderson** today. This brings the total US membership to 2,487 and the number of international members to 336.

Election to the National Academy of Engineering is among the highest professional distinctions accorded to an engineer. Academy membership honors those who have made outstanding contributions in at least one of the following categories: "engineering practice, research, or education," "pioneering of new and developing fields of technology, major advancements in traditional fields of engineering, or development/implementation of innovative approaches to engineering education," or "engineering leadership of one or more major endeavors." NAE members are among the world's most accomplished engineers from business, academia, and government.

Founded in 1964, the NAE provides independent, objective analysis and advice to the nation, offering leadership and insights for a complex world. Elected by their peers, the ballot for this class was set in December and the final vote took place in January.

Individuals in the newly elected class will be formally inducted during the NAE's Annual Meeting on October 5, 2025. A list of the new members and international members follows, with their primary affiliations at the time of election and a brief statement of their principal engineering accomplishments.

New Members

Alex Acero, AI venture partner, Cadenza, Monte Sereno, CA. For leadership in research and development of spoken language technology.

Samuel Achilefu, professor and Lyda Hill Distinguished University Chair, Department of Biomedical Engineering, University of Texas

Southwestern Medical Center, Dallas. For contributions and international leadership in the molecular imaging of cancer and molecular nanotherapeutics using biophotonics and nuclear technologies.

Khalil Amine, Argonne Distinguished Fellow, Argonne National Laboratory, Lemont, IL. For leadership of developments in the materials science of batteries and energy storage devices.

Marsha Anderson Bomar, president elect, American Society of Civil Engineers, Duluth, GA. For advancing sustainable passenger and freight ground transportation systems and land use development.

Alexander Apostolov, principal engineer, OMICRON Electronics, Los Angeles. For leadership in the digital transformation of protection and control in electrical power systems.

Kyriacos Athanasiou, distinguished professor and Henry Samueli Chair, Biomedical Engineering,

University of California, Irvine. For contributions to the understanding and treatment of musculoskeletal afflictions and for leadership in bioengineering.

Paramvir Bahl, technical fellow and chief technology officer, Azure for Operators, Microsoft Corp., Bellevue, WA. For contributions to wireless networking, indoor localization, and edge computing, and for leadership in the mobile computing community.

Gerard Baillely, senior vice president, Corporate Research and Development, Procter & Gamble, Mason, OH. For leadership in creating innovative consumer packaging with enhanced performance, value, consumer satisfaction, and sustainability.

Oswaldo A. Bascur, principal, OSB Digital LLC, The Woodlands, TX. For applying phenomenological modeling techniques in mineral and processing plant performance optimization.

Martin Zdenek Bazant, E.G. Roos (1944) Chair Professor, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge. For contributions to nonlinear electrochemical and electrokinetic phenomena, including induced charge electroosmosis, shock electrodialysis, capacitive desalination, and energy storage applications.

Gregg T. Beckham, senior research fellow, Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO. For leadership in biological and chemical catalysis for the valorization and creation of materials from waste plastics and biomass.

Moshe E. Ben-Akiva, Edmund K. Turner Professor, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge. For advances in

transportation and infrastructure systems modeling and demand analysis.

Olester Benson, corporate scientist (retired), 3M Corporate Research Process Laboratory, Woodbury, MN. For advancing microreplication technologies, leading to the commercialization of reflective safety films and brightness-enhancing films for electronics.

Jessica Bian, vice president, Grid Services, Grid-X Partners, Acton, MA. For leadership and contributions in electric power system reliability performance metrics, and for advancement of power flow modeling and control.

Rodney Bowersox, associate dean for research and Ford I Professor, Department of Aerospace Engineering, Texas A&M University, College Station. For leadership and contributions in hypersonic aerodynamics and aerothermodynamics through the development of experimental facilities and diagnostic techniques.

Lisa Brannon-Peppas, founder and president, PeppChem Consulting, Austin, TX. For contributions to the targeted drug delivery of chemotherapeutic agents and contributions to biomaterials swelling, and for biomedical leadership.

Susan L. Brantley, Evan Pugh University Professor Emeritus and Atherton Professor, Department of Geosciences, Pennsylvania State University, University Park. For elucidating chemical, biological, and physical processes associated with the circulation of aqueous fluids in shallow hydrogeologic settings.

Robert C. Brown, Anson Marston Distinguished Professor and Gary and Donna Hoover Chair, Iowa State University, Ames. For engineering processes for controlled oxidation of lignocellulosic biomass and plas-

tic wastes into low-carbon-footprint energy, fuels, and chemicals.

Douglas C. Cameron, co-president and director, First Green Partners, Erie, CO. For driving the chimerical development of a synthetic, biology-based, environmentally friendly bioprocess for biofuels, food security, and industrial chemicals.

Pei Cao, vice president of engineering, YouTube's Trust & Safety, YouTube, San Bruno, CA. For contributions to web caching, search engine efficiency, and information quality.

Patrick Chapman, chief executive officer, Startup in Energy Systems, Austin, TX. For the development of reliable inverters enabling large-scale photovoltaic energy system deployment.

Tze-Chiang Chen, IBM Fellow and vice president of science and technology, IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY. For technical contributions and leadership in materials development and the implementation of logic and memory technologies.

Long-Qing Chen, Donald W. Hamer Professor, Department of Materials Science and Engineering, Pennsylvania State University, University Park. For making the phase-field method the most powerful tool for predicting the mesoscopic microstructure and properties of engineering materials.

Igor Cherepinsky, director, Sikorsky Innovations, Sikorsky, A Lockheed Martin Co., Stratford, CT. For technical achievement and leadership in the development of autonomous systems for vertical takeoff and landing (VTOL) aircraft.

Kwok Cheung, founder and chief executive officer, KC-Exousia Consulting, Redmond, WA. For contributions to the design and

implementation of energy market management systems in power grid control centers worldwide.

Jennifer R. Cochran, Addie and Al Macovski Professor and senior associate vice provost for research, Shriram Center for Bioengineering and Chemical Engineering, Bioengineering Department, Stanford University, Stanford, CA. For biopharmaceutical protein discovery and development, biotechnology entrepreneurship, and leadership in academic bioengineering.

Robert Cohen, president, Digital, Robotics, and Enabling Technology, Stryker, Weston, FL. For design contributions to orthopedic surgical procedures and devices and for additive manufacturing leadership.

Victoria Coleman, chief executive officer and head, North America Research & Technology, Acubed, Sunnyvale, CA. For innovations in software operations related to national and cyber security and consumer electronics.

Dorin Comaniciu, senior vice president, Artificial Intelligence and Digital Innovation, Siemens Healthineers, Princeton, NJ. For contributions to diagnostic imaging and image-guided therapy, leading to better diagnosis and treatment for numerous patients.

Charles L. Cooney, professor emeritus, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge. For contributions to biochemical and pharmaceutical manufacturing that propelled the establishment and growth of the global biotechnology industry.

Charles Crouse, principal engineer, AECOM, Seattle. For contributions to public safety and infrastructure resilience through leadership in transferring research advances into seismic design practice.

Alexa Dembek, chief technology and sustainability officer, DuPont, Wilmington, DE. For leadership in the development of light-weight protective fiber technologies and value creation through innovation, product stewardship, and sustainability.

Nance K. Dicciani, co-founder and chief executive officer, RTM Vital Signs LLC, Ambler, PA. For technical leadership driving innovation in medical devices, materials, and chemical processes.

Scott Diddams, professor and Robert H. Davis Endowed Chair, Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder, Boulder. For contributions to optical frequency combs and their applications.

Christopher R. Doerr, chief executive officer, Aloe Semiconductor Inc., Middletown, NJ. For contributions to integrated photonic circuits for telecommunications.

James H. Dooley, chief technology officer, Forest Concepts LLC, Auburn, WA. For engineering achievements in the design of innovative wood products, biomass processes, and processing equipment.

Martin Eberhard, entrepreneur (retired), Tiveni Inc., Friday Harbor, WA. For leadership in the development of electric vehicles and entrepreneurship in transportation.

Julie Eng, chief technology officer, Coherent Corp., Los Altos, CA. For contributions to high-speed optoelectronic devices and modules.

Arthur Guy Erdman, Richard C. Jordan Professor and director, Earl E. Bakken Medical Devices Center, University of Minnesota, Minneapolis. For contributions to kinematics, mechanical design education, and applications for machines and products.

David Estell, fellow and laureate, International Flavors & Fragrances

Inc., Palo Alto, CA. For the development of protein engineering technologies that resulted in products for human health, animal health, food, and personal care.

Judith L. Estrin, chief executive officer, J Labs LLC, Menlo Park, CA. For contributions to the early development and adoption of networking technologies, products, and services, and for leadership that advanced innovation.

Lorne G. Everett, chief scientist and chief executive officer, L. Everett & Associates LLC, Santa Barbara, CA. For establishing EPA-endorsed monitoring and cleanup methodologies for contaminated soil and groundwater.

Antonio Facchetti, professor and Hightower Chair, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta. For contributions to the engineering of commercially viable electronic materials and devices.

Thea Feyereisen, distinguished technical fellow, Honeywell Aerospace Technologies, Hudson, WI. For cockpit innovations, including synthetic vision systems, runway overrun awareness, and moving maps.

Yoel Fink, professor, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge. For the design and production of structured photonic fibers, enabling surgeries and the invention of fabrics that sense and communicate.

Dan M. Frangopol, Fazlur R. Khan Endowed Chair of Structural Engineering and Architecture, ATLSS Engineering Research Center, Lehigh University, Bethlehem, PA. For contributions to life-cycle civil engineering and leadership in its global development and adoption.

Marcelo H. García, professor and M.T. Geoffrey Yeh Endowed Chair, Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana. For research on sediment transport and river and marine morphodynamics, and for advancing hydraulics and sedimentation engineering.

Suresh V. Garimella, president and distinguished professor, University of Arizona, Tucson. For contributions to microscale heat and mass transport, academic leadership, and service to the nation.

Marian Gindy, vice president, Small Molecule Science and Technology, Manufacturing Division, Merck, Rahway, NJ. For pharmaceutical product innovation and for transforming new technology concepts into robust platforms, processes, and products.

Justin Hanes, Lewis J. Ort Professor, Wilmer Eye Institute, Johns Hopkins University, Baltimore. For developing innovative technologies that improve drug and gene delivery, resulting in multiple approved products.

Elizabeth Hausler, founder and chief executive officer, Build Change, Denver. For transformational impact as an international social entrepreneur, saving lives by building sustainable communities with natural hazard resilience.

Robert W. Heath, Charles Lee Powell Chair, Jacobs School of Engineering, University of California, San Diego, La Jolla. For contributions to the theory and practice of wireless communication.

Hooshang Heshmat, co-founder, president, and chief executive officer/technical director, Mohawk Innovative Technology Inc., Au Sable Forks, NY. For contributions to tribology, oil-free turbomachinery, and gas foil bearings.

Marillyn A. Hewson, chairman, president, and chief executive officer (retired), Lockheed Martin Corp., Bethesda, MD. For the development and delivery of breakthrough innovations and technology programs of critical importance to global security and space exploration.

Stephanie C. Hill, president, Rotary and Mission Systems, Lockheed Martin Corp., Cockeysville, MD. For leadership of aerospace and defense systems in cyber security, rotary wing systems, and unmanned aerial vehicles.

Elizabeth Ann Holm, Richard F. and Eleanor A. Towner Professor and department chair, Department of Materials Science and Engineering, University of Michigan, Ann Arbor. For achievements in the development and application of computational methods for understanding microstructural evolution and integrated computational materials engineering.

Philip John Holmes, professor emeritus, Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ. For the application of methods of nonlinear analysis and chaotic dynamics to engineering systems.

Charles W. Hull, co-founder and chief technology officer, 3D Systems Corp., Valencia, CA. For the invention of 3D printing and the subsequent development of the additive manufacturing industry.

Joseph G. Jacangelo, vice president and director of research, StanTec, Washington, DC. For advancing low-pressure membrane filtration for potable water treatment and wastewater reuse.

Andrew Jassy, president and chief executive officer, Amazon, Seattle. For driving the cloud computing revolution by envisioning, launch-

ing, leading, and growing Amazon Web Services.

Russell Taylor Johns, professor and George E. Trimble Chair, Department of Energy and Minerals Engineering, Pennsylvania State University, University Park. For developing thermodynamically rigorous algorithms for modeling hydrocarbon recovery processes.

Wayne R. Johnson, researcher, Ames Research Center, NASA, Moffett Field, CA. For contributions to rotorcraft comprehensive analysis, tiltrotor aircraft development, emerging electric aircraft, and the Mars Helicopter development.

Denise C. Johnson, group president, Caterpillar Inc., Irving, TX. For overseeing the development of zero-emission and autonomous mining equipment.

Abhay Joshi, president and chief executive officer, Axent Biosciences, San Mateo, CA. For taking botulinum toxin medicines from initial lab-scale production to global commercialization and distribution.

Mohan Kalkunte, vice president of architecture, Core Switch Products Group, Broadcom Limited, Palo Alto, CA. For contributions to ethernet switching for data networks.

Lydia E. Kavraki, Kenneth and Audrey Kennedy Professor, Department of Computer Science, Rice University, Houston. For developing randomized motion-planning algorithms for robotics and robotics-inspired methods in biomedicine.

Dale E. Klein, Reese Endowed Professor, Walker Department of Mechanical Engineering, University of Texas at Austin, Austin. For leadership in the development of programs that enhance the safety of nuclear power plants worldwide.

Nicholas A. Kotov, Joseph B. and Florence V. Cejka Professor, Univer-

sity of Michigan, Ann Arbor. For scalable methods for the self-assembly of anisotropic nanoparticles and multifunctional nanocomposites that led to commercially viable advanced materials.

Michael Krames, president, Arkesso LLC, Kentfield, CA. For contributions to high-brightness, visible-spectrum light-emitting diodes.

Ray Kreig, president, R.A. Kreig & Associates, Anchorage, Alaska. For developing the geotechnical engineering system used for the site selection and design of the Trans-Alaska Pipeline.

Concetta La Marca, senior principal consultant, The Chemours Co., Wilmington, DE. For contributions to the commercial development and manufacturing of advanced, non-ozone-depleting, low-global-warming-potential refrigerants.

Robert Lee, board member, Ene-dym, Brighton, MI. For leadership in automotive propulsion systems for conventional internal combustion hybrid and electric vehicles.

Maria C. Lehman, infrastructure market leader, GHD Group, Orchard Park, NY. For leadership in public service and business engineering management focused on the sustainability and resilience of critical civil infrastructure.

Donald Mark Lipkin, professor, Department of Materials Science and Engineering, Texas A&M University, College Station. For innovations of coatings and rhenium recovery that contributed to higher efficiency and sustainability of jet engines and gas turbines.

Michal Lipson, Eugene Higgins Professor, Department of Electrical Engineering and Applied Physics, Columbia University, New York City. For the development, fabrication, and implementation of build-

ing blocks of silicon photonics.

Yueh-Lin Loo, Theodora D. '78 and William H. Walton III '74 Professor, Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ. For contributions in developing processing-structure-property relationships in organic, polymer, and hybrid electronic materials, and leadership in decarbonizing shipping.

Tomás Lozano-Pérez, professor, Computer Science & Engineering, Massachusetts Institute of Technology, Cambridge. For contributions to robot motion planning and molecular design.

Subbaiah Malladi, chief technical officer and principal engineer, Exponent, Menlo Park, CA. For contributions to critical national-transportation-safety-related investigations and safety regulation enforcement.

John L. Manferdelli, confidential computing incubation leader (retired), VMware, San Francisco. For contributions and leadership in the field of trusted and confidential computing.

Christopher David Manning, Thomas M. Siebel Professor, Departments of Linguistics and Computer Science, Stanford University, Stanford, CA. For the development and dissemination of natural language processing methods.

Maja J. Matarić, Chan Soon-Shiong Chair and distinguished professor, Computer Science, Neuroscience, and Pediatrics, University of Southern California, Los Angeles. For contributions to human-robot interaction and socially assistive robotics.

Karen McKee, president, ExxonMobil Product Solutions Co., Spring, TX. For leadership and innovation in growing sustainable global petrochemical businesses.

William Mickols, president, Mickols Consulting LLC, Tenino, WA. For the discovery, development, and implementation of next-generation desalination membranes and related water filtration systems.

Leslie A. Momoda, executive vice president, HRL Laboratories LLC, Malibu, CA. For delivering materials innovation and application of advanced technologies across diverse industry sectors, ranging from hypersonics to high performance electronics.

Radhakrishnan Nagarajan, senior vice president and chief technology officer, Optical Platforms, Marvell Technology, Cupertino, CA. For advances in high-speed lasers and photonic integration technologies.

Alison Nordt, director, Space Science and Instrumentation, Advanced Technology Center, Lockheed Martin Corp., Palo Alto, CA. For contributions to aerospace and space science through development of novel instruments and space missions, extending understanding of the universe.

Aydogan Ozcan, Chancellor's Professor and Volgenau Chair, Department of Electrical and Computer Engineering and Bioengineering, University of California, Los Angeles. For contributions to mobile sensing and telepathology for medical diagnostics.

Mark Papermaster, chief technology officer and executive vice president, Advanced Micro Devices, Horseshoe Bay, TX. For leadership in the design and production of complex integrated circuit processors.

Beth L. Parker, professor and NSERC Industrial Research Chair, School of Engineering, University of Guelph, Canada. For developing field methods to monitor, understand, and remediate groundwater

contaminant occurrence, fate, and transport in fractured rock aquifers.

John Paserba, president and chief executive officer, Southern States LLC, Hampton, GA. For advancing the application of power electronic controllers to improve power system performance and grid reliability.

Fang Peng, RK Mellon Endowed Chair Professor, Swanson School of Engineering, University of Pittsburgh. For contributions to the development of high-power electronics technologies for advanced power grid control and energy conversion.

Kristin A. Persson, Daniel M. Tellep Distinguished Professor, Department of Materials Science and Engineering, University of California, Berkeley. For pioneering data-driven materials design through the creation and stewardship of open materials databases and associated data-mining algorithms.

Kristala L. Prather, Arthur D. Little Professor and department head, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge. For the development of innovative approaches to regulate metabolic flux in engineered microorganisms with applications to specialty chemicals production.

Madhusudan Raghavan, senior tech fellow and lab group manager, Battery and System Architecture Group, General Motors, West Bloomfield, MI. For contributions to kinematic analysis applied to conventional, hybrid, and electric vehicle powertrain designs.

Ellen M. Rathje, Janet S. Cockrell Centennial Chair, Maseeh Department of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, Austin. For contributions to seismic slope assessment and site response analysis and the development of cyberinfrastructure

for natural hazards engineering.

Eric Rignot, Donald Bren Professor, School of Physical Sciences, University of California, Irvine. For contributions to modeling and analysis of the effects of global climate change on polar ice sheets using radar interferometry.

Mary M. Roybal, senior principal engineering fellow (retired), Raytheon, Tucson, AZ. For contributions to the structural integrity and producibility of some of the world's most advanced missile systems.

Roger Ruan, professor and director, Department of Bioproducts and Biosystems Engineering, and Center for Biorefining, University of Minnesota, St. Paul. For scientific and entrepreneurial contributions to the sustainable development of the circular economy within food and agricultural systems.

Stuart J. Russell, professor, Division of Computer Science, University of California, Berkeley. For contributions to developments in artificial intelligence, including reasoning, probabilistic modeling, planning, safety, and education.

Richard Sandstrom, self-employed (retired), Sandstrom LLC, Encinitas, CA. For contributions and leadership in the development of commercial quality microlithography EUV lasers, enabling nanoscale chip manufacturing.

Mahadev Satyanarayanan, Jaime Carbonell University Professor, Computer Science Department, Carnegie Mellon University, Pittsburgh. For contributions to distributed file systems, edge computing, and mobile computing.

Linda S. Schadler, dean, College of Engineering and Mathematical Sciences, University of Vermont, Burlington. For contributions to the fundamental understanding, property control, and commercial appli-

cation of polymer nanocomposites.

David V. Schaffer, Hubbard Howe Jr. Distinguished Professor of Biochemical Engineering, College of Chemistry, University of California, Berkeley. For the application of fundamental molecular and cellular engineering principles to enable the clinical success of gene and cell therapies.

Hanspeter Schaub, distinguished professor, Colorado Center for Astrodynamics Research, University of Colorado Boulder, Boulder. For contributions to the control of satellite formations and relative orientations utilizing natural forces, including the use of electrostatics.

SawTeen See, president, See Robertson Structural Engineers LLC, Foster City, CA. For leadership in managing the innovative structural design of tall, super-tall, and other buildings.

Vladimir M. Shalaev, Robert and Anne Burnett Distinguished Professor, Electrical and Computer Engineering, Purdue University, West Lafayette, IN. For contributions to negative-refractive-index material at optical frequencies and metamaterial applications for national security.

Robert Smilowitz, senior principal, Protective Design & Security, Thornton Tomasetti, New York City. For protecting lives from acts of terrorism through vulnerability assessment, threat mitigation, and building standards development.

Alice E. Smith, Joe W. Forehand Jr. Distinguished Professor, Department of Industrial and Systems Engineering, Auburn University, Auburn, AL. For advancements in computational intelligence as applied to modeling and optimization of complex systems.

Richard W. Spinrad, professor emeritus, Oregon State University, Bend. For leadership in the development and execution of national and international programs in ocean sci-

ence and technology.

Eric Swanson, research affiliate, Massachusetts Institute of Technology, Gloucester. For contributions and entrepreneurship in biomedical imaging and optical communications.

Don Tan, fellow and chief engineer, Power Products, Northrop Grumman Corp., Redondo Beach, CA. For contributions to advances in power electronics technologies for space applications.

Marc Taub, vice president, Technical Operations, Abbott Diabetes Care, Alameda, CA. For leadership in bioengineering design and the development of transformational wearable devices and systems for in vivo electrochemical biosensing.

Thomas E. Taylor, chief engineer and technical director (retired), large processing systems, Northrop Grumman Corp., San Jose, CA. For technical leadership and innovations in the development of intelligence collection systems of great value to the United States.

Margaret Torn, earth senior scientist and program head, Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA. For contributions to the understanding of soil carbon dynamics and sustained leadership of the long-term monitoring of climate change.

Janeen Uzzell, chief executive officer, National Society of Black Engineers, Alexandria, VA. For providing access to health information and healthcare outcomes globally, and increasing racial, gender, and community engagement in engineering practice.

John Charles Vassberg, chief design officer, JetZero, Long Beach, CA. For advancing aircraft capabilities and efficiency through inventions, developments, and applications of aerodynamic technologies,

computational methods, and experimental validations.

Venkat Venkatasubramanian, Samuel Ruben-Peter G. Viele Professor, Columbia University, New York City. For research in the development and implementation of artificial intelligence methods in process safety and pharmaceutical manufacturing.

Evelyn N. Wang, Ford Professor, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge. For contributions to clean energy, water technology, and nanostructure-based phase change heat transfer, and for service to the nation.

Yong Wang, Regents Professor and Voiland Distinguished Professor in Chemical Engineering, School of Chemical Engineering and Bioengineering, Washington State University; and laboratory fellow and associate director of the Institute for Integrated Catalysis, Pacific Northwest National Laboratory and Washington State University, Pullman. For advancing catalyst discovery, design, and reaction engineering for energy and environmental applications.

Zhen-Gang Wang, Dick and Barbara Dickinson Professor, Department of Chemical Engineering, California Institute of Technology, Pasadena. For unifying theories of thermodynamic and transport properties of polymers to predict phase behavior, self-assembly, and nucleation.

Robert A. Weller, senior scientist, Physical Oceanography, Woods Hole Oceanographic Institution, Falmouth, MA. For contributions to the theory, observation, and instrumentation of the upper ocean and air-sea interface.

Jim Wiethorn, founder and chairman, International Crane & Construction Safety Solutions LLC,

Sugar Land, TX. For research on and analysis of crane accidents, saving lives through improved technology and training.

Changfu Wu, expert mechanical engineer and senior scientific reviewer, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD. For propelling the United States to a position of global leadership in early patient access to cutting-edge cardiovascular devices.

Yuehwen Yih, Tompkins Professor, Edwardson School of Industrial Engineering, Purdue University, West Lafayette, IN. For contributions to supply chain management systems in humanitarian relief efforts and health care.

Eric S. Yuan, founder and chief executive officer, Zoom Video Communications Inc., San Jose, CA. For leadership enabling worldwide video communications.

Lixia Zhang, Jonathan B. Postel Chair, Computer Science Department, University of California, Los Angeles. For the development of internet protocols which significantly impacted internet quality and performance, and for leadership in setting standards.

Ming Zhou, chief engineer, Computational Mechanics and Design Optimization, Altair Engineering Inc., Newport Beach, CA. For contributions to topology optimization and simulation-driven generative design technology.

Julie Zimmerman, vice provost for planetary solutions and professor, School of the Environment, Yale University, New Haven, CT. For leadership in education and the development of green technologies that enhance the sustainability of engineered systems.

New International Members

Hirofumi Akagi, distinguished professor, Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan. For contributions to the theory, design, and application of utility high-power electronic systems.

Mohammed Y. Al Qahtani, downstream president, Aramco, Dhahran, Saudi Arabia. For leadership and contributions in carbon management in oil and gas operations.

Virgilio A.F. Almeida, professor emeritus, Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil. For contributions to the performance evaluation and modeling of large-scale distributed systems and to information technology policies.

Vanderlei Salvador Bagnato, professor, Institute of Physics, University of São Paulo, São Carlos, Brazil. For contributions to metrology, optical sensing, bioengineering, technology transfer, and engineering education.

Catia Bastioli, chief executive officer, Novamont, Novara, Italy. For leadership in industrial bioplastics and for championing the circular bioeconomy.

Jon Atli Benediktsson, rector/president and professor, Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik. For contributions to design and application of advanced machine learning and morphological methods in information analysis of multi-source sensing data.

Martin Blunt, chair, Department of Earth Science and Engineering, Imperial College London, London, United Kingdom. For the description and simulation of multiphase flow in porous media from pore-to-reservoir-scale.

Anil Chaudhry, managing director and zone president, Schneider Electric India Foundation, Gurugram. For leadership in advancing real-time power grid management.

Bjerne S. Clausen, president and chief executive officer (retired), Topsoe A/S, Hørsholm, Denmark. For pioneering the molecular science of catalyst function for important chemical processes with long-lasting technical and practical impact.

Daniel Gold, head, DDR&D (MAFAT), Israel Ministry of Defense, Tel Aviv. For contributions and leadership in the defense industry, especially the development and deployment of missile defense systems.

Masanori Hamada, professor emeritus, Department of Civil Engineering, Waseda University, Yokohama City, Japan. For developing methods to protect critical infrastructure against earthquakes and soil liquefaction.

Gerhard A. Holzapfel, professor, Institute of Biomechanics, Graz University of Technology, Graz, Austria. For contributions to the mechanics of soft materials and for the quantification, modeling, and analysis of biological tissue microstructures.

Katalin Karikó, adjunct professor, University of Pennsylvania, Jenkintown. For the development of nucleoside-modified mRNA used in the world's leading COVID-19 lipid nanoparticle vaccines.

Ingrid Kögel-Knabner, chair, Department of Soil Science, Technical University of Munich, Freising, Germany. For contributions to mechanisms for soil organic carbon persistence and impacts on soil functions and climate.

Fumio Koyama, professor emeritus, Tokyo Institute of Technology, Yokohama, Japan. For developing vertical-cavity surface-emitting lasers

(VCSEL), including the first room-temperature continuous operation and advanced beam-steering VCSEL.

Ondrej Ladislav Krivánek, president, Nion Co., Kirkland, WA. For the realization of the aberration-corrected scanning transmission electron microscope to extend microscopy to sub-ångström resolution.

Hitoshi Kuninaka, director general, Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagami-hara. For accomplishments in electric propulsion for spacecraft and for the Hayabusa2 asteroid sample return mission.

Po-Yen Lu, partner, CID Group, Taipei, Taiwan. For leadership and innovation in the mass production of liquid-crystal displays and semiconductors for high-bandwidth data transmission.

Dierk Raabe, director, Department of Microstructure Physics and Alloy Design, Max Planck Institute for Sustainable Materials, Düsseldorf, Germany. For contributions to the fundamental understanding of materials and to sustainable metallurgy.

Bhaskar Ramamurthi, professor, Department of Electrical Engineering, Indian Institute of Technology, Madras, Chennai, India. For the advancement and commercialization of wireless technology in India.

Mate Rimac, founder and chief executive officer, Rimac Group, Sveta Nedelja, Croatia. For contributions to electric vehicle technology and leadership in the automotive industry.

Sreedhara Panicker Somanath, former secretary and chairman, Department of Space, Indian Space Research Organisation, Bangalore, India. For engineering excellence and leadership in making India the first country to reach the Moon's south pole, advancing space technology.

NAE Newsmakers

Gilda Barabino (NAM), president, Olin College of Engineering, received the distinction of **Corresponding Academician of the Academy of Sciences of Cuba** (ACC). She was honored in September 2024 during “Opportunities for U.S.-Cuba-Mexico Collaboration in Public Health,” held in Mexico City and organized by AAAS, ACC, and two Mexican universities.

Richard G. Baraniuk, C. Sidney Burrus Professor of Electrical and Computer Engineering, Rice University, has been awarded the **2025 IEEE Jack S. Kilby Signal Processing Medal**. Dr. Baraniuk is recognized for his contributions to multiscale and sparse signal processing.

Regina Barzilay (NAM), distinguished professor for AI and health, Massachusetts Institute of Technology, has been chosen to receive the **IEEE 2025 Frances E. Allen Medal** for innovative machine learning algorithms that have led to advances in human language technology and demonstrated impact on the field of medicine.

Marsha J. Berger (NAS), professor of computer science and mathematics, New York University, has been awarded the **2025 John von Neumann Prize**, the highest honor and flagship lecture of the Society for Industrial and Applied Mathematics, in recognition of her foundational work in adaptive mesh refinement and embedded boundary methods for partial differential equations.

James J. Collins (NAS/NAM), Henry Termeer Professor of Medical Engineering & Science, Massachusetts Institute of Technology, earned the **2025 IEEE Medal for Innovations in Healthcare Technology** for his work in synthetic gene circuits

and programmable cells, launching the field of synthetic biology, and impacting healthcare applications.

Nicholas M. Donofrio, CEO, NMD Consulting LLC, was awarded the **National Action Council for Minorities in Engineering 2024 Percy Pierre Lifetime Achievement Award** for his persistent long-term leadership and support of diversity, equity, and inclusion in the engineering disciplines.

Russell D. Dupuis, Steve W. Chadwick Endowed Professor Emeritus, Georgia Institute of Technology, is the recipient of the **2025 Japan Prize** honoring achievements in the fields of materials science and production. Professor Dupuis was chosen for development of metalorganic chemical vapor deposition technology for compound semiconductor electronic and optoelectronic devices, and pioneering contribution to its large-scale commercialization.

Andrea J. Goldsmith, dean, School of Engineering and Applied Science, Princeton University, is the recipient of the **IEEE 2025 Mildred Dresselhaus Medal** for contributions to and leadership in wireless communications theory and practice. Dr. Goldsmith was also the recipient of the **2024 IEEE James H. Mulligan Jr. Award** for educating, mentoring, and inspiring generations of students, and for authoring pioneering textbooks in advanced digital communications.

Richard D. Gitlin, distinguished university professor, emeritus, University of South Florida, received the **IEEE 2025 Alexander Graham Bell Medal** for high-impact contributions to communications systems and networks.

Joseph P. Heremans, Ohio Eminent Scholar and professor, and **Longya Xu**, distinguished professor of engineering, Ohio State University, have been elected members of the **National Academy of Inventors 2024 Class of Fellows**.

Takeo Kanade, U.A. and Helen Whitaker University Professor of Computer Science and Robotics, Carnegie Mellon University; **R. Vijay Kumar**, professor and Nemirovsky Family Dean of the School of Engineering and Applied Science, University of Pennsylvania; and **Daniela Rus** (NAS), director of the MIT Computer Science and Artificial Intelligence Laboratory, will share the **2024 John Scott Award** presented by the Board of Directors of City Trusts of Philadelphia. The three were honored for their research that has revolutionized the science of robotics while dramatically improving the quality of human life. This prestigious award that honors the scientific legacy of Benjamin Franklin was presented in Philadelphia on November 20, 2024. Dr. Rus has also been awarded the IEEE’s **2025 Edison Medal** for sustained leadership and pioneering contributions in modern robotics. This follows her **IEEE 2023 Robotics and Automation Award** for pioneering contributions to the design, realization, and theoretical foundations of innovative distributed, networked autonomous systems.

David L. Kaplan, Stern Family Professor of Engineering and distinguished university professor, Tufts University, has been named **one of the world’s most highly cited researchers**. Researchers on the Clarivate 2024 list have a significant

impact on the research community as judged by the rate their work is cited by their peers. Only about 1 in 1,000 researchers worldwide qualify. Dr. Kaplan's research focus is on biopolymer engineering, tissue engineering, regenerative medicine, and cellular agriculture.

Ahsan Kareem, Robert M. Moran Professor of Engineering, University of Notre Dame, has been elected to the **European Academy of Sciences and Arts**. He will join the academy's division of technical and environmental sciences.

Dina Katabi (NAS), Thuan & Nicole Pham Professor, Massachusetts Institute of Technology, received the **2023 Kuwait Prize in Emerging Specialized Sciences** from the Kuwait Foundation for the Advancement of Science. The award celebrates her groundbreaking work bridging computer science and electrical engineering, significantly advancing the speed, reliability, and security of data transfer in wi-fi and cellular systems.

Oussama Khatib, professor of computer science and director of Stanford University's Robotics Lab, has won the **Great Arab Minds Award in Engineering and Technology**. Professor Khatib's achievements include groundbreaking research and innovations in robotic systems, algorithms, and sensing technologies, which have significantly advanced the field, enabling robots to operate effectively in diverse environments, including healthcare, industrial automation, and deep-sea exploration.

Rob Knight, director of the Center for Microbiome Innovation, University of California, San Diego, has been named the **2025 Scientist of the Year** by the Achievement Rewards for College Students (ARCS) Foundation San Diego

chapter. An international leader in the scientific study of microbes, Dr. Knight will be honored on April 13 at a special event hosted by the ARCS Foundation.

Jennifer A. Lewis (NAS), Hansjorg Wyss Professor of Biologically Inspired Engineering, Harvard University, has won the **2025 National Academy of Sciences James Prize in Science and Technology Integration** for pioneering contributions to the programmable assembly of soft functional, structural, and biological materials.

Asad M. Madni, independent consultant and retired president, COO, and CTO, BEI Technologies Inc., has been selected for induction in the **2025 class of the Space Technology Hall of Fame®** for the GyroChip®. According to STHF, "this prestigious recognition honors the significant impact your technology has made by bringing space innovations to benefit life on Earth. Your technology stood out among a competitive field of nominees, and we are thrilled to celebrate your achievement."

Thomas L. Marzetta, distinguished industry professor, New York University Tandon School of Engineering, received the **2025 IEEE Eric E. Sumner Award** for originating the Massive MIMO technology in wireless communications. The award recognizes outstanding contributions to communications technology.

Kiran Mazumdar-Shaw, executive chairperson, Biocon Limited, has been conferred the **Jamsetji Tata Award** by the Indian Society for Quality. This lifetime achievement award was presented to her for pioneering the biosciences movement in India and serving patients in India and across the world through an unwavering focus on quality.

Nicholas W. McKeown, professor of electrical engineering and computer science, Stanford University, is the recipient of the **2025 Marconi Prize**. Professor McKeown is recognized for his fundamental contributions to high-performance switches and routers and to software defined networking and for transferring these contributions into widespread practice.

José M.F. Moura, Philip & Marsha Dowd University Professor, Carnegie Mellon University, received the **2025 IEEE Haraden Pratt Award** for sustained leadership and outstanding contributions to the IEEE in education, technical activities, awards, and global connections. The award is given to an IEEE senior member or fellow for outstanding volunteer service to IEEE.

Dan Peer, professor and director, Lab of Precision NanoMedicine, Tel Aviv University, was elected a **fellow of the United States National Academy of Inventors**.

Roderic Ivan Pettigrew (NAM), vice chancellor for health and strategic initiatives, Texas A&M University System, was awarded the **2024 Research Achievement Award** by the American Heart Association. AHA noted that the "type of research Roderic is conducting takes ideas that only seemed to exist in science fiction and makes them a reality. The potential for these new biomedical inventions to help people maintain good health longer and his role merging engineering and medicine will help lead to advancements that shape the future of medicine." The award was presented during the Presidential Session of the AHA's Scientific Sessions 2024 in Chicago.

Marc Raibert, CEO and founder, Boston Dynamics, has won the **IEEE**

2025 Robotics and Automation Award for pioneering and leading the field of dynamic legged locomotion.

Henry Samuelli, chairman of the board, Broadcom Corporation, is the **IEEE 2025 Medal of Honor Laureate** and the first recipient of its \$2 million monetary prize. IEEE bestowed its highest award on him in recognition of his pioneering research and advances in developing and commercializing analog and mixed signal communications systems circuits, which led to the explosive growth of broadband communication, wired and wireless networking technologies used today. Dr. Samuelli is also honored for his significant philanthropic support for science, technology, engineering, and math (STEM) education.

Lisa T. Su, president and CEO, Advanced Micro Devices Inc., has been named **2024 CEO of the Year** by *Time* magazine. The recognition highlights her transformative leadership that has propelled AMD into a dominant player in the semiconductor industry.

Levi T. Thompson, Elizabeth Inez Kelley Professor of Chemical Engineering, University of Delaware, was awarded a **2024 Medal of Distinction** by the university. Professor Thompson, who served as dean of the College of Engineering from 2018 until 2024, was recognized for his service and leadership.

Stephen Wei-Lun Tsai, research professor emeritus, Stanford University, will receive the **2025 Daniel**

Guggenheim Medal during AIAA's 2025 Awards Gala on April 30. Dr. Tsai was chosen for his series of pioneering innovations revolutionizing design and simplifying manufacturing processes of composites, shaping today's composites industry.

Chris G. Van de Walle, distinguished professor of materials and Herbert Kroemer Chair in Materials Science, University of California, Santa Barbara, has been awarded the American Physical Society's **2025 Aneesur Rahman Prize for Computational Physics**, the highest honor given by APS for work in that field. Professor Van de Walle was chosen in recognition of his development and application of first-principles methods for computing the structural, electronic, and optoelectronic properties of point defects and interfaces.

Vladimir N. Vapnik, professor, Columbia University, along with Ekaterina Zhuravskaya of the Paris School of Economics, has received the **G.A. Gamow Award** of the Russian-American Association of Scientists. The award is given to encourage members of the Russian-speaking scientific diaspora for outstanding achievements recognized by the wider scientific community. Professor Vapnik received the honor for his fundamental contributions to statistical learning theory and, especially, development of a widely used approach known as Support Vector Machines that formed the foundation of modern data science.

Ganapati D. Yadav, National Science Chair (Government of India), chairman of the board of governors at LIT University, and the former vice chancellor of the Institute of Chemical Technology, Mumbai, has been awarded the **Bhatnagar Fellowship**. The fellowship, given by the Council of Scientific and Industrial Research, is awarded annually for notable and outstanding research, applied or fundamental. It enables outstanding scientists to pursue excellence in scientific research and innovative technology development. Professor Yadav has also been selected for the **Eminent Engineer Award** by the Engineering Council of India (ECI) for his contributions to research and consultancy. He will be accorded the award in Delhi on April 28.

The **2025 Queen Elizabeth Prize for Engineering** has been awarded to seven engineers who have made seminal contributions to the development of modern machine learning, a core component of artificial intelligence advancements. NAE members awarded are **William J. Dally**, chief scientist and senior vice president of research, Nvidia Corporation; **Geoffrey E. Hinton** (NAS), emeritus professor, University of Toronto; **Jen-Hsun Huang**, CEO, Nvidia Corporation; **Yann A. LeCun** (NAS), chief AI scientist, Facebook Professor, New York University; and **Fei-Fei Li** (NAM), Sequoia Capital Professor Operations, Information, and Technology, Stanford University.

2025 Fritz J. and Dolores H. Russ Prize Acceptance Remarks by Ian Shanks

President Anderson, Vice President Leo, members of the Russ family and the NAE, ladies and gentlemen—I would like, most sincerely, to thank the National Academy of Engineering and Ohio University for their roles in administering and funding the Fritz J. and Dolores H. Russ Prize in Bioengineering and for honoring me with this most prestigious award from this august international organization. This award is in recognition of my invention of the enzyme biosensor, called the electrochemical capillary fill device (eCFD), and I am delighted to accept this top international award in bioengineering. This marks the first time it has been awarded to a UK engineer. I thank each and every one of you for being here on this auspicious occasion. I feel truly humbled.

I also wish to thank the Russ family for the vision and generosity of Fritz and Dolores in creating and supporting the prize and helping plug the gap left by the absence of a Nobel Prize for Engineering.

I would particularly like to acknowledge and thank Cyril Hilsum, a brilliant international member of the NAE, for nominating me for this signal honor and, also, my supporters who provided their personal assessments of both me and the impact of my invention.

¹ The 2025 Fritz J. and Dolores H. Russ Prize was presented to Ian Alexander Shanks, vice president of physical and engineering sciences (retired), Unilever, “for the invention of the electrochemical capillary fill device (eCFD), which gives diabetes patients and caregivers accurate and timely blood glucose measurements for diabetes management.”



Donald Leo, executive vice president and provost, Ohio University; Ian Alexander Shanks, 2025 Fritz J. and Dolores H. Russ Prize winner; and John Anderson, NAE president.

I also wish to thank “Young, Deborah”—as she appears many times in my inbox—for her advice and guidance on getting me here on time and in good order and for supervising the dinner, the paperwork, the presentation, and much else. Thank you, Deborah, and to your colleagues and contacts! Finally, I would like to acknowledge the help, tolerance, and support provided by Janice, my wife, without which I would not be here today. Thank you, Janice!

Like Fritz, I also came from a working family. A Scot from modest beginnings, I graduated with a BSc in electrical engineering from the University of Glasgow and went on to work for my government, conduct-

ing research in the then little-known field of liquid crystals.

Simplicity and cost were at the core of my invention of the eCFD. Having conducted pioneering research on liquid crystal displays (LCDs) for 10 years, I moved to Unilever to initiate research in biosensors—a field that combined the devices of physics with the molecules of biology. This “out of the box” thinking resulted in an innovation that integrated the physical sciences with engineering and biology.

I was able to learn enough about the latter to draw on my expert knowledge of LCDs to envision how the mass manufacture of digital watch LCD displays could be adapted to create many billions of inexpensive,

accurate, and simple-to-use blood glucose biosensor test strips that could be used by diabetics, easily, rapidly, and virtually painlessly, to self-test their own blood glucose levels and avoid the often-dire consequences of their disease.

Although Unilever patented the eCFD, it chose not to exploit the invention, but after 12 years of simply keeping it in force, they granted fully paid-up licenses, on request, for the eCFD to be manufactured and sold by a wide range of medical product companies around the world, including many based here in the United States. This underpinned the creation of a world market with sales measured in hundreds of billions of

dollars and has enabled hundreds of millions of diabetics globally to monitor and control their affliction.

There is also a legal dimension to this story. In 2019, after 13 years of financially and reputationally perilous litigation, the UK Supreme Court, in a landmark decision, awarded me substantial compensation from Unilever, as a fair share of their outstanding benefit from licensing the eCFD patents. This also established legal precedent for how the UK 1977 Patents Act should be interpreted in the future, thus inspiring and incentivizing future UK employee inventors. This was the first such award in over 40 years.

I'm pleased that the diabetes test

strip I invented still remains, in a variety of proprietary forms and about 30 years later, the gold standard for controlling the disease and for checking the calibration of the new, wearable, continuous glucose monitor products now on the market.

Slightly paraphrasing Sir Winston Churchill's dictum on the characteristics of a good speech, I have tried to make these remarks short enough to be interesting but long enough to cover the essentials!

Thank you again for being here and listening to me, and thank you again to the NAE for awarding me this outstanding prize.

Engineering Innovation for a Brighter Future: The 2025 NAE National Meeting

Engineers, researchers, students, and emerging leaders gathered last month to participate in an expansive conversation about the future of engineering at the National Academy of Engineering's (NAE's) National Meeting that took place Feb. 12–14 at the Beckman Center in Irvine, California. Over three days, the meeting featured lectures from early career engineers, a keynote presentation by Nobel Laureate and NAE member Dr. **Frances H. Arnold**, and a panel discussion exploring the grand challenges of energy, medicine, space exploration, and quantum technologies.

The event kicked off with a dinner celebrating significant contributions in support of engineering. Dr. **John L. Anderson** reflected on his time as president of the NAE, and several donors were celebrated for their generous support of the NAE.

Dr. **Shu Chien** was awarded a Franklin medal; Dr. **Dianne Chong** and Dr. **John Wall** were awarded Curie medals; Dr. **Efi Foufoula-Georgiou**, Dr. Tryphon Georgiou, and Dr. Christian Lutkemeyer were awarded Einstein medals; and a Heritage medal was presented to Dr. **Fiona Doyle**.

The meeting opened with the dedication of the Beckman Center's Board Room to Anderson, who has served as NAE's president since 2019 and will conclude his term on June 30, 2025. NAE Council members and speakers gathered to celebrate Anderson's legacy as NAE Vice President **Wes Harris** reflected on his impact and contributions to strengthening NAE's mission.

Following the dedication, Anderson and NAE Executive Officer Dr. **Alton Romig** officially opened the meeting, *Engineering Innovation for a Brighter Future*. This year's theme

highlighted the importance of ingenuity in engineering and of taking creative risks to solve modern challenges. The agenda was designed to inspire and encourage the next generation of engineers to embrace innovation, learn from failures, and shape the future of the field.

The Gilbreth Lectures

The Gilbreth Lectures, a highlight of the event, showcased four early career engineers selected through NAE's Grainger Foundation Frontiers of Engineering program.² Their talks, attended by high school students from Samueli Academy in Santa Ana, California, and Cabrillo Engineering and Design High School in Long Beach, California, explored cutting-edge research in healthcare, artificial intelligence (AI), sustainability, and advanced technology.

² <https://www.naefrontiers.org/16100.aspx>

Dr. Shriya Srinivasan, a professor of bioengineering at the Harvard School of Engineering and Applied Sciences, discussed her work on a new class of technologies to address gut health — ingestible devices that have the potential to transform understanding and treatment of gut disorders. (Video of this presentation is unavailable.) Srinivasan described her approach to designing neural interfaces that are better able to navigate the enteric nervous system, which acts as a kind of control center for digestion. She presented several examples, including Robo-Cap, an orally ingested capsule that breaks down the gut's mucus barrier, allowing better absorption of drugs that typically must be injected. Srinivasan highlighted the promise of ingestible devices, which can help us better understand and monitor internal organs and make treatment more accessible and personalized, a breakthrough comparable to wearable devices when they emerged.

The next speaker, Dr. Amanda Randles, a professor of biomedical engineering at Duke University, presented³ her vision for innovative solutions to intractable problems in healthcare. Rather than treating illnesses reactively, Randles' research seeks to identify and address diseases before symptoms even appear. Her work focuses on developing vascular digital twins — virtual models that can replicate patients' vasculature to create a visual representation of blood flow that can proactively show where problems may occur. These digital twins use advanced algorithms to simulate different care approaches, helping patients and doctors make

informed decisions before a condition worsens and enabling individuals to make lifestyle decisions based on personalized predictive models.

Sustainability and public health concerns were recurring themes throughout the lectures. Dr. Ishi Keenum, a professor of environmental engineering at Michigan Technological University, spoke about a growing issue: antibiotic-resistant bacteria in wastewater. Although wastewater treatment systems are highly effective at cleaning non-potable water, some resistant bacteria remain, creating potential public health concerns around water reuse. Keenum's research focuses on improving wastewater treatment to eliminate antimicrobial resistant bacteria; however, her work has been challenged by inconsistent measurements, which stem from differences in the methods used across labs. Her presentation highlighted the need to develop better models for comparability in environmental metagenomics.

The conversation on technological advancement continued with Dr. Angel Yanguas-Gil, a principal materials scientist at Argonne National Laboratory, who introduced a new approach to smart technology — one inspired by the insect brain. Current AI models, while impressive, require extensive training before deployment and then cannot adapt or learn further. Yanguas-Gil's research focuses on developing smart sensors that continuously learn while in use, taking inspiration from biological intelligence — brains that are power-efficient, adaptable, and capable of multitasking. Rather than modeling AI after the human brain, which is vastly complex, Yanguas-Gil pointed to insects as the perfect starting point. Their compact yet highly efficient learning systems could revolutionize

the design of AI-driven sensors and semiconductors, making them more resilient and adaptive while requiring less energy.

The meeting also provided valuable networking opportunities for students. Between sessions and during receptions, students had the chance to engage one-on-one with NAE members and speakers, discussing engineering topics and career aspirations in depth. Some students return year after year to stay informed about the latest advancements in science and technology, while for others, the NAE's National Meeting is a gateway into the world of engineering. For many, the meeting is a rare opportunity to meet professionals from disciplines they may never have encountered otherwise. These conversations offer invaluable insight and inspiration, helping shape the career paths of future engineers.

Inspiring the Next Generation of Engineers

On Friday, the inaugural Early Career Engineering Convocation (E2C2) took place. This meeting was introduced this year to foster enduring interactions between NAE members and early career engineers from academia, national labs, and industry. Frances H. Arnold gave a fireside chat, sharing her journey as an engineer and the unexpected turns that shaped her career. She reflected on how dynamic the field of engineering is, noting that, while she started as a chemical engineer, her path eventually led her to entrepreneurship and even got her a cameo on *The Big Bang Theory*.

Arnold emphasized the importance of adaptability and cross-disciplinary thinking, explaining that engineers don't need to be experts in every field to draw valuable insights

³ Recordings of 2025 NAE National Meeting presentations are available at the NAE Youtube Channel (<https://www.youtube.com/@theNAEng>).

from them. She encouraged early career engineers to embrace trial and error as a fundamental part of the process, rather than fearing mistakes. Some of the most groundbreaking discoveries, Arnold said, come from pushing past uncertainty and exploring ideas beyond one's immediate expertise.

Her discussion set the tone for the Early Career Engineering Convocation, which included panel discussions and breakout sessions covering energy innovation, medical advancements, space exploration, and quantum computing.

The panel discussion, moderated by engineer and TV host Deanne Bell, brought together experts from various fields to discuss the next generation of engineering solutions. The speakers explored how their disciplines — AI data centers, energy systems, medical technology, and space exploration — are evolving to become more efficient and cost-effective.

They also shared some of the most exciting developments in their respective fields and how they envision engineering shaping the future.

Dr. José N. Reyes Jr. spoke about the growing energy demands of AI data centers and the need for more efficient ways to power them. As AI technology continues to expand, he said, finding sustainable energy solutions will be crucial to ensuring long-term efficiency.

Dr. Michael A. Mussallem discussed how the latest advancements in medical technology — from innovative surgical tools to next-generation medical devices — are transforming patient care and improving health outcomes.

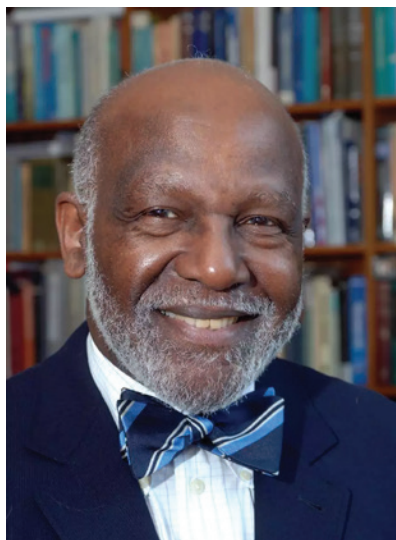
Dr. Laurie Leshin explored the latest developments in space exploration and how engineering continues to push the boundaries of what's possible beyond Earth. She spoke about NASA's ongoing missions and the role of private space companies in

accelerating innovation.

Dr. Tsu-Jae King Liu covered the future of quantum computing, making the case that as quantum computing progresses, it has the potential to revolutionize industries ranging from cybersecurity to drug discovery. She explained that new advancements in semiconductor technology are bringing us closer to a new era of computational power.

Following the panel, attendees separated into smaller, focused breakout groups, where they had the opportunity to engage in in-depth, cross-disciplinary discussions before reconvening for a debrief session, where representatives from each group shared key takeaways and proposed action items. For many of the young engineers and students in attendance, this meeting could be the spark that inspires them to pursue careers in engineering — tackling challenges and solving problems that today we can only begin to imagine.

Message from NAE Vice President Wesley L. Harris



I am pleased to share the 2024 fundraising update, highlighting another remarkable year of generosity and commitment from our NAE members and friends as we celebrated a milestone—our 60th anniversary and continued focus on the NAE's comprehensive fundraising effort, the *NAE Campaign for Leadership in a World of Accelerating Change*.⁴ I am deeply grateful for your continued commitment to the NAE. Thanks to your dedication, we raised over \$8.3 million in new gifts, pledges, and planned gifts in 2024, bringing our total campaign contributions to an impressive \$75.9 million. This prog-

ress is a testament to the trust you place in the NAE, and we remain deeply grateful for your ongoing investment in the NAE's mission. As we reflect on the past six decades of impact, it's inspiring to see how far the NAE has come in advancing engineering excellence, and your contributions make this work possible.

In honor of our 60th anniversary, we set ambitious fundraising goals as part of our "60 for the 60th" initiative. I'm proud to announce that we have successfully met or made significant progress on these goals—securing 60 new Curie, Franklin, or Lincoln Donor Society members, establishing 50 new named or endowed

⁴ <https://www.nae.edu/acceleratingchange/>

funds, and achieving a 51% member giving participation rate in our campaign. Thanks to your support, we also reached our Legacy Challenge objectives by surpassing the target of 60 new planned gifts, raising a total of \$3 million, thus bringing our total expectancies to \$10.6 million in future funding. These milestones for the Legacy Challenge reflect the outstanding leadership of Corale L. Brierley, Heritage Society Chair, and Elana Lippa, Director of Planned Giving, whose efforts have been crucial in making these achievements possible.

I am also delighted to announce that the NAE Council once again achieved 100% giving participation for the 10th consecutive year in 2024. This continued commitment is a testament to the Council's leadership and dedication to our mission. With their support, we are aiming for an 11th consecutive year of 100% giving participation in 2025.

As part of the continued success of our campaign, the NAE welcomed two new Lincoln Society members, four new Franklin Society members, six new Curie Society members, 12 new Einstein Society members, and a whopping 31 new Golden Bridge Society members to our Lifetime Giv-

ing Societies in 2024. I hope you will join me in welcoming these new Lifetime Giving Society members (See pages 83-91 for the full list). Your support ensures the NAE's future and empowers us to address pressing engineering challenges.

Looking ahead, we are eager to carry the momentum from 2024 into 2025. To ensure that we have adequate resources to cover the annual operations and maintenance of the Beckman Center, a fund of at least \$10 million is required to be properly capitalized. In 2024, we secured over \$3M for the National Academy of Engineering Foundation Beckman Center Endowed Fund to name the John L. Anderson Beckman Boardroom, which was unveiled and dedicated as part of the NAE National Meeting in February 2025. In 2025, The Grainger Foundation Frontiers of Engineering will also celebrate its 30th anniversary. We will also be welcoming a new NAE president and are excited to build on the strong foundation laid during President John L. Anderson's leadership over the last six years.

Philanthropic support continues to be vital to the NAE's success. With over 30% of members donating in 2024, we saw significant engagement

and support. Additionally, sections across the country exceeded their goals for the 60th Anniversary Giving Challenge, demonstrating the power of collective action and community within the NAE.

As we enter the final two years of our *Campaign for Leadership in a World of Accelerating Change*, I want to once again express my sincere thanks for your ongoing generosity. Your support is critical to advancing the NAE's mission to advance the welfare and prosperity of the nation by providing independent advice on matters involving engineering and technology, and by promoting a vibrant engineering profession and public appreciation of engineering.

If you are interested in learning more about how you can support the NAE or make a planned gift to ensure long-term sustainability, please don't hesitate to reach out to Radka Nebesky, Senior Director of Development at RNebesky@nae.edu or 202.207.8508. We remain grateful for your continued engagement and look forward to what we can accomplish together in the years ahead.

Sincerely,
Wesley L. Harris
NAE Vice President
(July 1, 2022–June 30, 2026)

2024 Honor Roll of Donors

Lifetime Giving Societies

We greatly appreciate the generosity of our donors. Your contributions enhance the impact of the National Academy of Engineering's work and support its vital role as a national advisor. The NAE acknowledges the following members and friends who have made generous charitable lifetime contributions. The gifts reflected on this list are as of December 31, 2024.

The Abraham Lincoln Society

In recognition of members and friends who have made lifetime contributions of \$1 million or more to the National Academy of Sciences, National Academy of Engineering, or National Academy of Medicine. Boldfaced names are NAE members.

Bruce and Betty Alberts	Mary E. Wilson	John F. McDonnell	Fritz J. and Dolores
Richard and Rita*	Bernard M. Gordon	Robin K. and Rose M.	H. Russ Prize Fund
Atkinson	Cecil H. Green*	McGuire	of the Russ College
Norman R. Augustine	John O. and Candace E.	George P. Mitchell*	of Engineering and
Craig and Barbara Barrett	Hallquist	The Ambrose Monell	Technology at Ohio
Jordan and Rhoda	Michael and Sheila Held*	Foundation	University
Baruch*	Jane E. Henney and Robert	The Moore Family	William J. Rutter
Stephen D. Bechtel Jr.*	Graham	Philip* and Sima	Bernard and Rhoda
Arnold and Mabel	William R. and Rosemary	Needleman	Sarnat*
Beckman*	B. Hewlett*	Peter O'Donnell, Jr.*	Leonard D. Schaeffer
Leonard Blavatnik	Ming and Eva Hsieh	Gilbert S. Omenn and	Sara Lee and Axel Schupf
Harry E. Bovay Jr.*	Irwin and Joan* Jacobs	Martha A. Darling	James H.* and Marilyn
Donald Bren	Robert L.* and Anne K.	Jonathan B. and Donna J.	Simons
George* and Virginia	James	Perlin	John and Janet Swanson
Bugliarello	Kenneth A. Jonsson*	Robert* and Mayari	Marci and James J.
Ralph J.* and Carol M.	Fred Kavli*	Pritzker	Truchard
Cicerone	Daniel E. Koshland Jr.*	Richard L. and Hinda G.	Tachi* and Leslie Yamada
Ross and Stephanie	Tillie K. Lubin*	Rosenthal*	Anthony J. Yun and
Corotis	Whitney* and Betty	Martine A. Rothblatt	Kimberly A. Bazar
Harvey V. Fineberg and	MacMillan	Jack W. and Valerie Rowe	Anonymous (1)

The Benjamin Franklin Society

In recognition of members and friends who have made lifetime contributions of \$500,000 to \$999,999 to the National Academy of Sciences, National Academy of Engineering, or National Academy of Medicine. Boldfaced names are NAE members.

John and Pat Anderson	Chau-Chyun and Li-Li	Penny and Bill George,	Alexander Hollaender*
Rose-Marie and Jack R.	Chen	George Family	Thomas V. Jones*
Anderson*	Shu and Kuang-Chung	Foundation	Cindy and Jeong Kim
John and Elizabeth	Chien	Christa and Detlef Gloge	Ralph and Claire Landau*
Armstrong	James McConnell Clark*	William T.* and Catherine	Asta and William W.
Kenneth E. Behring*	Barry and Bobbi Collier	Morrison Golden	Lang*
Gordon Bell*	Henry David*	Jerome H.* and Barbara N.	Marcia K. McNutt
Elkan R.* and Gail F.	Richard Evans*	Grossman	Ruben F.* and Donna
Blout	Eugene Garfield	Wesley L. Harris	Mettler
Russell L. Carson	Foundation	John L. Hennessy	Dane* and Mary Louise
Charina Endowment Fund	Theodore Geballe*	Lyda Hill Philanthropies	Miller

*Deceased

Oliver E. and Gerda K.
Nelson*
Shela and **Kumar Patel**
Richard F. and Terri W.
Rashid

Henry and Susan
Samueli
Herbert A. and Dorothea
P. Simon*

Raymond and Maria
Stata
Roy and Diana Vagelos
Andrew and Erna*
Viterbi

Alan M. Voorhees*
Anonymous (2)

The Marie Curie Society

In recognition of members and friends who have made lifetime contributions of \$250,000 to \$499,999 to the National Academy of Sciences, National Academy of Engineering, or National Academy of Medicine. Boldfaced names are NAE members.

The Agouron Institute
W.O. Baker*
Warren L. Batts
Elwyn* and Jennifer
Berlekamp
Gopa and **Arindam Bose**
Paul F. Boulous
Daniel and Lana Branton
Gail H. Cassell
Sigrid and **Vint Cerf**
Dianne Chong and David
Squiers
Clarence S. Coe*
Rosie and Stirling A.
Colgate*
W. Dale and Jeanne C.
Compton*
Ruth David and Stan
Dains
Lance and Susan* **Davis**
Nicholas M. Donofrio
David and Miriam
Donoho

Ruth and Victor Dzau
James O. Ellis, Jr. and
Elisabeth Paté-Cornell
Dotty* and **Gordon**
England
Edward H. Frank and
Sarah G. Ratchye
William L.* and Mary Kay
Friend
George and Christine
Gloeckler
Paul and Judy **Gray**
Diane Greene and **Mendel**
Rosenblum
Diane E. Griffin*
Robert A. Hicks*
Chad and Ann **Holliday**
William R. Jackson*
Anita K. Jones
Mary and **Howard Kehrl***
Jill Howell **Kramer**
Kent Kresa
Mark and Becky **Levin**

Frances and **George Ligler**
Azad M. and Carla C.
Madni
Stella and **Steve* Matson**
William W. McGuire
Janet and **Richard M.***
Morrow
Clayton Daniel and
Patricia L. **Mote**
Ralph S. O'Connor*
Kenneth H. Olsen*
Larry* and Carol **Papay**
Stephen Quake
Ann and **Michael Ramage**
Simon Ramo*
Anne and **Walt* Robb**
Matthew L. Rogers and
Swati Mylavarapu
Julie and **Alton D. Romig**
Jr.
Jonathan J. Rubinstein
Stephen and Anne **Ryan***
Wendy and **Eric Schmidt**

H.E. Simmons*
Charles and Lisa **Simonyi**
Pradeep Sindhu and
Marie Bertrand
Robert F. and Lee S.
Sproull
Edward C. Stone*
Judy Swanson
Ted Turner
Leslie L. Vadasz
Martha **Vaughan***
Charles M.* and Rebecca
M. Vest
Robert* and Robyn
Wagoner
John C. Wall
David Walt and Michele
May
Susan R. Wessler
Wm. A. Wulf*
Anonymous (2)

The Einstein Society

In recognition of members and friends who have made lifetime contributions of \$100,000 to \$249,999 to the National Academy of Sciences, National Academy of Engineering, or National Academy of Medicine. Boldfaced names are NAE members.

Nancy Adler*, Arnold
Milstein, and Julia
Adler-Milstein
Laura E. and John D.
Arnold
Holt Ashley*
Nadine Aubry and John L.
Batton
Francisco J.* and Hana
Ayala

William F. Ballhaus Sr.*
David Baltimore
Thomas D.* and Janice H.
Barrow
H.H. and Eleanor F.
Barschall*
Forest and Kathy **Baskett**
Sunanda Basu
Donald and Joan Beall
Daniel and Frances **Berg**

Diane and Norman
Bernstein*
Bharati and **Murty**
Bhavaraju
Chip and Belinda
Blankenship
Erich Bloch*
Barry W. Boehm*
David G. Bradley
Lewis M. Branscomb*

John and Sharon Brauman
Sydney Brenner*
Eric A. Brewer
Corale L. Brierley and
James A. Brierley
Lenore and **Rob Briskman**
Kristine L. Bueche
Malin Burnham
Ursula Burns and Lloyd
Bean*

*Deceased

Eugen and **Cleopatra Cabuz**
 Christine Cassel and Michael McCally
 John K. Castle
Josephine Cheng
 Priscilla and **Sunlin* Chou**
 Vinay and **Uma* Chowdhry**
John and **Assia Cioffi**
Paul Citron and Margaret Carlson Citron
A. James Clark*
 George W.* and Charlotte R. Clark
James H. Clark
 Ellen Wright Clayton and Jay Clayton
G. Wayne Clough
 Fred E. Cohen and Carolyn B. Klebanoff
 John D. Corbett*
 Pedro M. Cuatrecasas
 Roy Curtiss and Josephine* Clark-Curtiss
 Elaine and **Erroll Davis**
Jeffrey Dean and Heidi Hopper
Pablo G. and Silvia **Debenedetti**
Zhonghan John Deng
 Peter Dervan and Jacqueline Barton
 Roman W. DeSanctis
 Julie H. and Robert J. Desnick
Robert* and Florence **Deutsch**
 Paul M. Doty*
 Charles W. Duncan, Jr.*
 George and Maggie Eads
Robert and Cornelia **Eaton**
 Emanuel and Peggy Epstein*
 Thomas E. Everhart
 Peter Farrell
 Michiko So* and Lawrence Finegold

Tobie and **Daniel J.* Fink**
 Delbert A. and Beverly C. Fisher
George and **Ann Fisher**
Robert C. and Marilyn G. **Forney***
Harold K.* and Betty **Forsen**
Efi Foufoula-Georgiou and Tryphon Georgiou
 Kenneth R. Fulton
 Yellow H. Gandhi
Elsa M. Garmire and Robert H. Russell
 Melinda F. Gates
William H. Gates III
 Lou and Robin **Gerstner**
 Nan and **Chuck* Geschke**
 Jack and Linda Gill
 Sid Gilman and Carol Barbour
Martin E. and Lucinda **Glicksman**
 Avram Goldstein*
Robert W. Gore*
Susan L. Graham and Michael A. Harrison
Robert K. Grasselli* and Eva-Maria Hauck-Grasselli
 Corbin Gwaltney*
Carol K. Hall
 Margaret A. Hamburg and Peter F. Brown
 William M. Haney III
George* and Daphne **Hatsopoulos**
 Robert M. Hauser
 John G. Hildebrand and Gail D. Burd
 Jane Hirsh
 Jackie and **Freeman Hrabowski**
 Yijing and Nils Hultgren
 Scott J. Hultgren
Michael W. Hunkapiller
 Catherine Adams Hutt and Peter Barton Hutt
Jennie S. Hwang

M. Blakeman Ingle
 Kathleen Jamieson
 Trina and Michael Johns
 Richard B. Johnston, Jr.
Frank and Pam **Joklik**
Trevor O. Jones
Robert E. Kahn and Patrice A. Lyons
Thomas Kailath
Paul and Julie **Kaminski**
 Yuet Wai and Alvera Kan
John and Wilma **Kassakian**
Michael Kavanaugh and Carol James
 Diana S. and **Michael D. King**
Leon K. and Olga **Kirchmayer***
David B. Kirk
 Frederick A. Klingenstein*
 William I. Koch
 Gail F. Koshland
Wook Hyun Kwon and Son Cha Yi
John W. Landis*
 Janet and Barry Lang
 Louis Lange
 Ming-wai Lau
Gerald and Doris **Laubach**
Edward D. Lazowska and Lyndsay C. Downs
David M.* and Natalie **Lederman**
 Peter and Susan Leem
 Bonnie Berger and **Frank Thomson Leighton**
Thomas M. Leps*
 Jane and Norman N. Li
R. Noel Longuemare, Jr.
 Christian Lütkemeyer
Asad M., Gowhartaj, and Jamal **Madni**
 Davis L. Masten and Christopher Ireland
 Jane and **Roger L. McCarthy**
 Michael and Pat McGinnis

Burt* and Deedee **McMurtry**
Jan and **D. Nathan Meehan**
 Rahul Mehta
Richard and Martha **Meserve**
Robert and Robyn **Metcalfe**
 G. William* and Ariadna Miller
 Ronald D. Miller
 Stanley L. Miller*
James K.* and Holly T. **Mitchell**
Sanjit K. and Nandita **Mitra**
 Sharon and **Arthur Money**
 Joe and Glenna Moore
 David* and Lindsay Morgenthaler
 G. Michael Morris
 Cherry A. Murray
Narayana and Sudha **Murty**
 Jaya and Venky **Narayanamurti**
 Ellen and **Philip Neches**
 Norman F. Ness*
Ronald and Joan **Nordgren**
 Susan and **Franklin M. Orr, Jr.**
David Packard*
Roberto Padovani
Charles and Doris **Pankow***
Jack S. Parker*
 Nirmala and Arogyaswami **J. Paulraj**
 Cathy Peercy
 Edward E. Penhoet
Julia M. Phillips and John A. Connor
Percy A. and Olga* **Pierre**
Allen E.* and Marilyn **Puckett**
 Alexander Rich*
 Arthur D. Riggs*

*Deceased

Emanuel P. Rivers
Ronald L. Rivest
 Eugene and Ruth
 Roberts*
Ann K. Roby and
 Lawrence Dodd
 Christopher Rogers
Howie Rosen and Susan
 Doherty
 Henry M. Rowan*
Joseph E. and Anne P.
 Rowe*
 John M. Samuels, Jr.
Linda S. Sanford
Maxine L. Savitz
 Charles Sawyers
 Randy Schekman and
 Sabeeha Merchant

Walter Schlup*
 Nicole Shanahan
Heung-Yeung (Harry)
Shum and Ka Yan
Richard P. Simmons
 Harold C. and Carol H.
 Sox
 Georges C. St. Laurent,
 Jr.
Arnold and Constance
Stancell
Richard J. and Bobby
 Ann Stegemeier
 F. William Studier
Lisa T. Su
Michael Albert and
 Elizabeth Ann **Sutton**
 Thomas and Marilyn

Sutton
 Charlotte and **Morris***
Tanenbaum
Sridhar Tayur and
 Gunjan Kedia
Peter* and Vivian **Teets**
Hemant K. and Suniti
Thapar
 Samuel O. Thier
Leonard Kent* and
 Kayleen **Thomas**
James M. Tien and Ellen
 S. Weston
Gary and Diane **Tooker**
 Katherine K. and **John J.**
Tracy
 Holly and **Jeff Ullman**
 Patricia Bray-Ward and

David C.* Ward
 James N. Weinstein
Robert* and Joan
Wertheim
Robert M.* and Mavis E.
White
 John C. Whitehead*
 Jean D. Wilson*
Ken Xie
 Tachi* and Leslie Yamada
 Tilahun D. Yilma
Yannis and Sheryl **Yortsos**
Adrian Zaccaria*
Alejandro Zaffaroni*
 Peter Zandan
Elias A. Zerhouni
 Janet and Jerry Zucker
 Anonymous (3)

Golden Bridge Society

In recognition of NAE members and friends who have made lifetime contributions totaling \$20,000 to \$99,999. Bold-faced names are NAE members.

\$75,000 to \$99,999

Johanna M.H. Levelt
Sengers*

John Neerhout, Jr.
Buddy Ratner and Cheryl
 Cromer

Robert E.* and Mary L.
Schafrik

Donald and Linda **Winter**

\$50,000 to \$74,999

Andreas* and Juana
 Acrivos
Frances H. Arnold
 Jane K. and **William F.**
Ballhaus, Jr.
Becky and Tom **Bergman**
Mark T. Bohr
Rudy and Anna
Bonaparte
Corbett Caudill
William Cavanaugh
Selim A. Chacour
Joseph M. Colucci
 The Crown Family
Fiona M. Doyle and
 Stephen C. Blair
Gerard W. Elverum
Bruce and Pat **Finlayson**

Edith M. Flanigen
Samuel C. Florman*
Douglas W. and Margaret
 P. **Fuerstenau**
 Elizabeth and **Paul H.***
Gilbert
Joseph W. Goodman
Priscilla and **Paul E.* Gray**
Kathy and **Albert**
Greenberg
Eli Harari
Mark Horowitz
Bahman Hoveida
Evelyn L. Hu and **David L.**
Clarke
James R.* and Isabelle
Katzer
Terri L. Kelly

Kathryn S. and **Peter S.**
Kim
Louis J. and M. Yvonne
 DeWolf **Lanzerotti**
David C. Larbalestier
Kai Li
Larry V. McIntire
Nicholas W. McKeown
Charles W. Moorman
Darla and **George E.***
Mueller
 Jane and **Alan Mulally**
 Cynthia J. and **Norman**
A.* Nadel
Robert M.* and Marilyn
R. Nerem
 Matt O'Donnell
David and Linda **Patterson**

Fernando C. Pereira
 Srilatha and **Prabhakar**
Raghavan
 Ellen and **George A.***
Roberts
Jerry Sanders III
 Donna and **Jan Schilling**
Richard Szeliski
Leo John* and Joanne
Thomas
David W. Thompson
Suzanne M. Vautrinot
Darsh T. Wasan
Sheila E. Widnall
Taiyin Yang
A. Thomas Young
William and Sherry **Young**
Charles F. Zukoski

*Deceased

\$20,000 to \$49,999

Linda M. Abriola and Larry M. Albert Ilesanmi and Patience Adesdia Rodney C. Adkins Jane E. and Ronald J. Adrian Alice Merner Agogino Ann and Monty Alger Vickie G. and David J. Allstot Valerie and William A. Anders John C. Angus* Seta and Diran Apelian Ruth and Ken Arnold Kamla* and Bishnu S. Atal Wanda M. and Wade Austin Arthur B. and Carol Baggeroer Ruzena K. Bajcsy Clyde* and Jeanette Baker William F. Baker William F. Banholzer David K. Barton* Steven Battel Diane and Samuel W.* Bodman Anjan and Francy Bose Kathleen and H. Kent Bowen Andrei Z. Broder Rodney A. Brooks Alan C. Brown* Andrew and Malaney L. Brown Robert L. Byer Federico Capasso and Paola Salvini-Capasso François J. Castaing* Lili Cheng Joe H. and Doris W.L. Chow Margaret S.Y. Chu and Tze-Yao Chu	Harry M. Conger Red Conger and Sissy Sailors Kay and Gary* Cowger Karen and Steve Crouch Natalie W. Crawford Mary Czerwinski Glen T. and Patricia B. Daigger David and Susan Daniel Ingrid Daubechies and Robert Calderbank L. Berkley Davis Carl de Boor Aart de Geus and Esther John Mary and Raymond Decker Tom and Bettie* Deen Sevgi and Ali Dogru Elisabeth M. Drake E. Linn Draper, Jr. James J. Duderstadt Bonnie J. Dunbar Delores M. Etter Stephen N. Finger Anthony E. Fiorato G. David Forney, Jr. Stephen R. Forrest Alec D. Gallimore and Reates K. Curry Efi Foufoula-Georgiou and Tryphon Georgiou Katharine G. Frase and Kevin P. McAuliffe David A. Friedman and Paulette J. Meyer Richard L. and Lois E. Garwin Arthur and Helen Geoffrion Eduardo D. Glandt Arthur L. and Vida F. Goldstein Deborah L. Grubbe and James B. Porter, Jr. Hermann K. Gummel*	Ashraf Habibullah Joyce and James Harris Janina and Siegfried Hecker Chris T. and Kathleen* Hendrickson Robert and Darlene Hermann David* and Susan Hodges Grace and Thom Hodgson Urs Höelzle Lee Hood and Valerie Logan Hood John Howell and Susan Conway Yingzhi and Xuedong (XD) Huang J. Stuart Hunter Ray R. Irani Mary Jane and Vernon Irwin Wilhelmina and Stephen Jaffe Leah H. Jamieson Edward G.* and Naomi Jefferson James O. Jirsa Kristina M. Johnson Michael R. Johnson John L. and Nancy E. Junkins Eric W. and Karen F. Kaler Min H. Kao Brian W. Kernighan James L. Kirtley, Jr. Albert S. and Elizabeth M. Kobayashi Demetrious Koutsoftas James N. Krebs* Lester C.* and Joan M. Krogh David J. Kuck Ellen J. Kullman H.T. and Ling-Ling Kung Michael and Christine Ladisch	Lois and Butler Lampson Yoon-Woo Lee Frederick J. Leonberger Robert A. Lieberman Burn-Jeng Lin Jack E. Little Chen-Ching Liu and Hiromi Okumura Robert G. Loewy* Lester L. Lyles William J. MacKnight Thomas* and Caroline Maddock Guru Madhavan and Ramya Ramaswami Thomas J. Malone* Savita and Ajay Malshe Henrique S. Malvar David A. Markle W. Allen Marr David and Diane Matlock Robert D. Maurer Dan and Dalia* Maydan Jyoti* and Aparajita Mazumder Kiran Mazumdar-Shaw James C. McGroddy Kishor C. Mehta David G. Messerschmitt R.K. Michel James J. Mikulski Susan M. and Richard B. Miles Piotr D. Moncarz Duncan T. Moore Van and Barbara Mow Jagdish Narayan Paul and Dotty Nielsen William D. Nix Ellen Ochoa Thomas and Jo Overbye Claire L. Parkinson Ellen M. Pawlikowski John H. Perepezko Aliene and Thomas K. Perkins Lee* and Bill Perry
---	---	--	---

*Deceased

Donald E. Petersen Leonard and Diane Fineblum Pinchuk John W. and Susan M. Poduska Dana A. Powers William F. Powers Henry H. Rachford, Jr.* John F. and Diane Reid Kenneth and Martha Reifsnider Thomas J. Richardson Richard J.* and Bonnie B. Robbins Bernard I. Robertson Ronald S. Robinson Mary Ann and Thomas Romesser William B.* and Priscilla Russel Vinod K. Sahney Steve* and Kathryn Sample	William and Emily Sanders Jose G. Santiesteban Kathryn Sarpong Ronald V. Schmidt* John H. Schmertmann Fred B. Schneider and Mimi Bussan William R. Schowalter Lyle H. Schwartz and Celesta S. Jurkovich Sharon and Norman Scott Charles L. Seitz Surendra P. Shah Martin B. and Beatrice E.* Sherwin Jennifer Sinclair Curtis Larry L. Smarr Megan J. Smith Alfred Z. Spector and Rhonda G. Kost David B. and Virginia H.	Spencer Gunter Stein Charles L. Seitz Larry L. Smarr Ann and Ken Stinson Mary and Charles Sukup Virginia and Carl Sulzberger Yongkui Sun Gaye and Alan Taub Rosemary and George Tchobanglous Edwin L. Thomas Matthew V. Tirrell Jean W. Tom James A. Trainham and Linda D. Waters John R. Treichler Cody and Richard Truly David M. Van Wie Raymond Viskanta* Thomas H. and Dee M.	Vonder Haar Albert R.C.* and Jeannie Westwood David and Tilly Whelan George M. Whitesides Ward O. and Mary Jo Winer Edgar S. Woolard, Jr. Israel J. Wygnanski Beverly and Loring Wyllie Jie Xue Murty V.V.S. and Bhavani Yalla James A. and Joan E. Yurko Jie Zhang Teresa and Steve Zinkle Mary Lou and Mark D. Zoback Anonymous (3)
--	--	---	--

Heritage Society

In recognition of members and friends who have included the National Academy of Sciences, National Academy of Engineering, or National Academy of Medicine in their estate plans or who have made some other type of planned gift to the National Academies. Boldfaced names are NAE members.

H. Norman and Idelle Abramson* Bruce and Betty Alberts Gene M.* and Marian Amdahl Joan* and Edward Anders John C. Angus* John and Elizabeth Armstrong Norman R. Augustine Jack D. Barchas Harrison H. and Catherine C.* Barrett Stanley Baum* Clyde J. Behney Ruth Behrens C. Elisabeth Belmont Daniel and Frances Berg Paul Berg* Bobbie A. Berkowitz	Elkan R.* and Gail F. Blout Enriqueta C. Bond John and Ariel Bowers Daniel and Lana Branton Robert and Lillian Brent* Corale L. Brierley James A. Brierley Lenore and Rob Briskman Kristine L. Bueche George* and Virginia Bugliarello Dorit Carmelli Charles and Martha Casey Peggy and Thomas* Caskey Barbara and Don Chaffin A. Ray Chamberlain Linda and Frank Chisari Rita K. Chow	Paul Citron and Margaret Carlson Citron John A. Clements Morrel H. Cohen Stanley N. Cohen Graham A. Colditz and Patti L. Cox Ross and Stephanie Corotis Ellis* and Bettsy Cowling Barbara J. Culliton Glen T. and Patricia B. Daigger David and Susan Daniel Tom and Bettie* Deen Julie H. and Robert J. Desnick Peter N. Devreotes Fiona M. Doyle and Stephen C. Blair	Ruth and Victor Dzau Gerard W. Elverum Dotty* and Gordon England Emanuel and Peggy Epstein* Harvey V. Fineberg and Mary E. Wilson Tobie and Daniel J.* Fink Bruce and Pat Finlayson Robert C. and Marilyn G. Forney* William L.* and Mary Kay Friend Mitchell H. Gail and Ruth Pfeiffer Arthur and Helen Geoffrion Irma Gigli
---	--	--	--

*Deceased

Elizabeth and **Paul H.***
Gilbert
 Sid Gilman and Carol
 Barbour
Martin E. and Lucinda
Glicksman
 George and Christine
 Gloeckler
 Christa and Detlef Gloge
Joseph W. Goodman
 Chushiro* and Yoshiko
 Hayashi
 Jane Henney and Robert
 Graham
 Bradford H. Gray
 Joe W. Gray
 John G. Hildebrand and
 Gail D. Burd
John R. Howell and
 Susan Conway
 Peter M. Howley
 Catherine Adams Hutt
 and Peter Barton Hutt
 Nancy S. and Thomas S.
 Inui
 Anthony A. James and
 Cynthia K. French
 Richard B. Johnston, Jr.
Anita K. Jones
 Jerome Kagan*
 Michael Katz and Robin

J. Roy
Michael and Carol
Kavanaugh
 K. I. Kellermann
 Diana S. and **Michael D.**
King
 Michael D. Lairmore
 Norma M. Lang
 Brian A. Larkins
 Marigold Linton and
 Robert Barnhill
Daniel P. Loucks
 Ruth Watson Lubic
 R. Duncan* and Carolyn
 Scheer Luce
Thomas* and Caroline
Maddock
 Asad and Taj **Madni**
 Rudolph A. Marcus
 Pat and Jim McLaughlin
 Jane Menken
 Sharon and **Arthur**
Money
Van and Barbara **Mow**
 Mary O. Mundinger
 Philip* and Sima
 Needleman
 Norman F. Ness*
Ronald and Joan
Nordgren
 Godfrey P. Oakley

Gilbert S. Omenn and
 Martha A. Darling
Bradford W. and Virginia
W. Parkinson
Zack T. Pate*
 Neil and Barbara
 Pedersen
 Cathy Peercy
 Thomas and Patty Pollard
 Louis J. Ptacek
 James J. Reisa, Jr.
 Emanuel P. Rivers
Richard J.* and Bonnie
B. Robbins
 Eugene* and Ruth
 Roberts
 Julie and **Alton D.**
Romig, Jr.
James F. Roth*
 Esther and Lewis*
 Rowland
 Sheila A. Ryan
 Paul R. Schimmel
 Stuart F. Schlossman
Lyle H. Schwartz
 Rudi* and Sonja Schmid
 Vern L. and Deanna D.
 Schramm
 Susan C. Scrimshaw
 Kenneth I. Shine
Arnold and Constance

Stancell
 H. Eugene Stanley
 Harold Stark
 Rosemary A. Stevens
John and Janet **Swanson**
Esther Sans Takeuchi
 Paul and Pamela Talalay*
 Walter and Marti Unger
Robert* and Robyn
Wagoner
John C. Wall
 Patricia Bray-Ward and
 David C. Ward*
 Clare M. Waterman
 Myron and Linda
 Weisfeldt
Robert* and Joan
Wertheim
 C. Kern Wildenthal
 Gail and Robert Wilensky
Donald and Linda
Winter
 Richard Witter
 Maw-Kuen Wu
Wm. A. Wulf*
 Tilahun D. Yilma
 Michael and Leslee
 Zubkoff
 Anonymous (3)

Loyalty Society

In recognition of members and friends who have made gifts to the National Academies of Sciences, Engineering, and Medicine for at least 20 years. Boldfaced names are NAE members.

Francois M. Abboud
Andreas* and Juana
Acrivos
 Stephen Adler
 Dyanne D. Affonso
 Bruce and Betty Alberts
 Barbara W. Alpert
 Lawrence K. Altman
John and Pat **Anderson**
John and Elizabeth
Armstrong
 Joan K. Austin
Arthur B. and Carol
Baggeroer

Donald W. Bahr
 John R. Ball
 Jack D. Barchas
 William A. Bardeen
 Clyde F. Barker
 Jeremiah A. Barondess
 Angela Barron McBride
 Gordon Baym and
 Cathrine Blom
Zdenek P. Bazant
 Terry and Roger N.
 Beachy
 Marlene and **Georges**
Belfort

Gordon Bell
 Leslie Z. Benet
 Kenneth I. Berns
 Carla F. Berry
 Mina J. Bissell
 Marie McCormick and
 Robert Blendon
 Floyd E. Bloom
Jack L. Blumenthal
 Enriqueta C. Bond
 Richard J. Bonnie
Lillian C. Borrone
 Kathleen and **H. Kent**
Bowen

Thompson Bowles
 John and Sharon Brauman
 Paula A. Braveman
 W.F. Brinkman
 Joseph A. Buckwalter
 Kathleen Coen Buckwalter
 Kristine L. Bueche
Jack E. Buffington
George* and Virginia
Bugliarello
 Martin Bukovac
James D. Callen
 Webster and Jill Cavenee
 Barbara and **Don Chaffin**

*Deceased

- David R. and Jacklyn A.
Challoner
Rita K. Chow
Ralph J.* and Carol M.
Cicerone
George W.* and Charlotte
Clark
Michael and Adriana
Clegg
John A. Clements
Linda Hawes Clever
G. Wayne Clough
Marshall H. Cohen
Morrel H. Cohen
Mary Sue Coleman
Barry and Bobbi Collier
R. John Collier
Richard A. Conway
R. James and Beverly
Cook
Max D. Cooper
Nicholas R.* and Linda
A. Cozzarelli
Pedro M. Cuatrecasas
Roy Curtiss and
Josephine
Clark-Curtiss*
G. Brent and Sharon A.
Dalrymple
James E. Darnell, Jr.
Ruth David and Stan
Danis
Lance and Susan* **Davis**
Igor B. and Keiko O.
Dawid
Pablo G. and Silvia
Debenedetti
Mary and **Raymond**
Decker
Tom and Bettie* **Deen**
Roman W. DeSanctis
Julie H. and Robert J.
Desnick
Don E. Detmer
William H. Dietz
Salvatore DiMauro
Nicholas M. Donofrio
Irwin Dorros
Earl H. Dowell
- R. Gordon Douglas, Jr.
E. Linn Draper, Jr.
James J. Duderstadt
Lewis S. Edelheit
David and Lucy T.
Eisenberg
Robert and Ingrid
Eisenman
Gerard W. Elverum
Robert M. Epstein
W. G. Ernst
Caswell A. Evans, Jr.
John V. Evans
Thomas E. Everhart
Gary Felsenfeld
Stanley Fields
Harvey V. Fineberg and
Mary E. Wilson
Tobie and Daniel J.* Fink
Delbert A. and Beverly C.
Fisher
Susan Fiske
Edith M. Flanigen
Kent V. Flannery
Samuel C. Florman
G. David Forney, Jr.
Harold K.* and Betty
Forsen
T. Kenneth Fowler
Hans* and Verena
Frauenfelder
Carl Frieden
William L.* and Mary
Kay **Friend**
Douglas W. and Margaret
P. **Fuerstenau**
Fred H. Gage
Mitchell H. Gail
Theodore V. Galambos
Joseph G. Gall
Elisabeth Gantt
Martin Gellert
Louis V. Gerstner, Jr.
Nan and **Chuck*** **Geschke**
John P. Geyman
Elizabeth and **Paul H.***
Gilbert
David Ginsburg
David V. Goeddel
- Lewis R. Goldfrank
Joseph W. Goodman
Enoch Gordis
Bradford H. Gray
Shirley and Harry Gray
Margie and Larry A.
Green
Robert B. Griffiths
Paul F. Griner
Michael Grossman
Edward E. Hagenlocker
Philip C. Hanawalt
Wesley L. Harris
Stanley R. Hart
Robert Mason Hauser
Janina and **Siegfried**
Hecker
Donald R. Helinski
Adam Heller
John L. Hennessy
Jane Henney and Robert
Graham
Arthur H. Heuer
George J. Hirasaki
John P. Hirth
David* and Susan
Hodges
Frank Hole
Peter M. Howley
Sarah and Dan Hrdy
Barbara S. Hulka
Catherine Adams Hutt
and Peter Barton Hutt
Richard and Fleur Hynes
Nancy S. and Thomas S.
Inui
Robert L.* and Anne K.
James
Paul C. Jennings
James O. Jirsa
Donald L. Johnson
Frank and Pam **Joklik**
Anita K. Jones
Marshall G. Jones
Michael M. Kaback
Paul and Julie **Kaminski**
Melvin F. Kanninen
John and Wilma
Kassakian
- Samuel L. Katz and
Catherine* M. Wilfert
James R.* and Isabelle
Katzer
K.I. Kellermann
Charles F. Kennel
Judson and Jeanne **King**
Albert S. and Elizabeth
M. **Kobayashi**
Bill and Ann **Koros**
Andrew H. Knoll
Joanne Knopoff
Jill Howell Kramer
Edward A. and Kathryn
F. Kravitz
Norma M. Lang
James S. and Elinor G.A.
Langer
Louis J. and M. Yvonne
DeWolf **Lanzerotti**
Gerald and Doris
Laubach
Judith R. Lave
Cynthia and Robert
Lawrence
Marvin & Annette Lee
Foundation: David Lee,
Trustee
Anthony Leggett
Ellen Lehman
I. Robert Lehman
Margaret A. LeMone
Alan and Agnes Leshner
Johanna M.H. Levelt
Sengers*
Howard Leventhal
Gene E. Likens
Peter W. Likins
Nathan and Barbara
Liskov
Jack E. Little
Robert G. Loewy*
Jane Lubchenco and
Bruce Menge
J. Ross Macdonald
Anthony P. Mahowald
Donald C. Malins
Vincent T. Marchesi
Joyce Marcus

*Deceased

Rudolph A. Marcus
William F. Marcuson, III
Robert C. Marini
Margo P. Marshak
Ida M. Martinson
Douglas S. Massey
David and Diane Matlock
Robert D. Maurer
William C. Maurer
Roger O. McClellan
Marie McCormick and
Robert Blendon
Christopher F. McKee
Marcia K. McNutt
Richard A. Meserve
Louis H. Miller
James K.* and Holly T.
Mitchell
Duncan T. Moore
Peter B. Moore
A. Stephen Morse
Joel Moses*
John H. Moxley III
Sezaki K. Mtingwa
Darla and **George E.***
Mueller
Earll M. Murman
Elaine Nadler
Jaya and **Venky**
Narayanamurti
Philip and Sima
Needleman
John Neerhout, Jr.
Stuart O. Nelson
Liz and Ben Neufeld

Elena and Stuart
Nightingale
William D. Nix
Ronald and Joan
Nordgren
Charles P. O'Brien
Gilbert S. Omenn and
Martha A. Darling
Gordon H. Orians
Jeremiah P. Ostriker
Larry* and Carol Papay
Barbara H. Partee
Robert E. Patricelli
Aliene and **Thomas K.**
Perkins
Jeffrey L. Platt
Thomas and Patty Pollard
Deborah E. Powell
William F. Powers
Michael Prats
William H. Press and
Jeffrey Howell
Roberta and **Edwin**
Przybylowicz
Roy Radner* and
Charlotte Kuh
Eli Reshotko
Kenneth A. Ribet
Charles M. Rice III
Charles C. Richardson
Jerome G. Rivard
Bernard I. Robertson
Julie and **Alton D. Romig,**
Jr.
Martine F. Roussel

Erkki Ruoslahti
Vinod K. Sahney
Steven B.* and Kathryn B.
Sample
Linda S. Sanford
Maxine L. Savitz
R. Duncan* and Carolyn
Scheer Luce
Randy Schekman
Joseph E. Scherger
Gerold L. Schiebler
Richard Schoen and Doris
Fischer-Colbrie
William R. Schowalter
Sara Lee and Axel Schupf
Henry G. Schwartz
Lyle H. Schwartz
Mischa Schwartz
John H. Schwarz
Robert J. Serafin
F. Stan Settles
Larry J. Shapiro
Charles J. Sherr
Kenneth I. Shine
Stephen M. Shortell
Edward H. Shortliffe
Arnold H. Silver
Maxine F. Singer
Jack M. Sipress
Harold C. and Carol H.
Sox
Patricia G. Spear
Robert F. and Lee S.
Sproull
Georges C. St. Laurent, Jr.

Raymond S. Stata
Richard J. and Bobby Ann
Stegemeier
Gunter Stein
Joan A. Steitz
Rosemary A. Stevens
Kenneth H. Stokoe
Edward C. Stone
Richard G. Strauch
Lubert and Andrea Stryer
F. William Studier
James M. Symons
Charlotte and **Morris**
Tanenbaum
Samuel O. Thier
James M. Tien and Ellen
S. Weston
Maury Tigner
Neil E. Todreas
Roxanne and Karl K.*
Turekian
Michael S. Turner
Roy and Diana Vagelos
Harold Varmus
Charles M.* and Rebecca
M. Vest
Andrew and Erna* Viterbi
Peter K. Vogt
Peter and Josephine von
Hippel
Thomas H. and Dee M.
Vonder Haar
Kenneth E. Warner
Joan Wennstrom Bennett

*Deceased

Great Hall Society (Annual Giving Society)

The Great Hall Society, the National Academies' annual giving society, recognizes donors who make leadership-level gifts at two levels: the silver level, for annual gifts totaling \$5,000–\$9,999, and the gold level, for annual gifts totaling \$10,000 or more. Annual funding provides resources that support ongoing activities and outreach efforts while also enabling the National Academies to respond to emerging issues facing the nation and world in a timely manner every year. We gratefully acknowledge the following members and friends who joined the Great Hall Society in 2024. Boldfaced names are NAE members.

Gold Level (\$10,000+)

Dyanne D. Affonso
Alice Merner Agogino
 Bruce and Betty Alberts
John and Pat Anderson
 Nancy C. Andrews
John and Elizabeth
Armstrong
 Richard and Rita*
 Atkinson
 Jane K. and **William F.**
Ballhaus, Jr.
 Peter Dervan and
 Jacqueline Barton
 Donald and Joan Beall
 Thomas Becker*
 May R. Berenbaum
 Paul Berg*
 Rosina M. Bierbaum
 Ellen Boer
Rudy and Anna
Bonaparte
 Gopa and **Arindam Bose**
Paul F. Boulos
 John* and Sharon
 Brauman
 Christopher S. Bretherton
Eric A. Brewer
Corale and Jim Brierley
Wesley G. Bush
 Eugen and Cleopatra
Cabuz
 Sigrid and **Vint Cerf**
Chau-Chyun and Li-Li
Chen
Josephine Cheng
Shu and Kuang-Chung
Chien

Dianne Chong and David
 Squiers
 Priscilla Chou
 Vinay Chowdhry
 Steven Chu
 Ralph J.* and Carol M.
 Cicerone
Paul Citron and Margaret
 Carlson Citron
David D. Clark
 Ellen Wright Clayton and
 Jay Clayton
 John A. Clements*
 Fred E. Cohen and
 Carolyn B. Klebanoff
 Graham A. Colditz and
 Patti L. Cox
 Barry and Bobbi Collier
Harry M. Conger, IV
 R. James and Beverly
 Cook
 Ross and Stephanie
Corotis
 Roy Curtiss and
 *Josephine
 Clark-Curtiss
Ruth David and Stan
 Dains
 Elaine and **Erroll Davis**
Aart de Geus and Esther
 John
Jeffrey Dean and Heidi
 Hopper
Pablo and Silvia
Debenedetti
 Julie H. and Robert J.
 Desnick
 Sevgi and **Ali Dogru**
Nicholas M. Donofrio

David and Miriam
 Donoho
Fiona M. Doyle and
 Stephen C. Blair
 Bob and Ann Dynes
 Ruth and Victor Dzau
James O. Ellis, Jr. and
Elisabeth Paté-Cornell
 Dotty* and **Gordon**
England
Delores M. Etter
 Nina Fedoroff
 Harvey V. Fineberg and
 Mary E. Wilson
Efi Foufoula-Georgiou
 and Tryphon Georgiou
Edward H. Frank and
 Sarah G. Ratchye
David A. Friedman and
 Paulette J. Meyer
 Kenneth R. Fulton
Lou and Robin Gerstner
 Joseph L. Goldstein
Joseph W. Goodman
Roy W. Gould*
 Fan C. Graham
 Jane E. Henney and
 Robert Graham
Susan L. Graham and
 Michael A. Harrison
 E. Peter Greenberg
 Diane E. Griffin*
 Jerome H.* and Barbara
 N. Grossman
Ashraf Habibullah
Carol K. Hall
John O. and Candace E.
Hallquist
 Margaret A. Hamburg

and Peter F. Brown
 Kathleen M. Harris
Wesley L. Harris
 Caroline S. Harwood
 Robert K. Grasselli*
 and Eva-Maria
 Hauck-Grasselli
 Robert Mason Hauser
John L. Hennessy
 Jane Henney and Robert
 Graham
 John G. Hildebrand and
 Gail D. Burd
 Lyda Hill Philanthropies
 Brigid L. Hogan
Chad and Ann Holliday
Bahman Hoveida
 Jackie and **Freeman**
Hrabowski
 Sarah and Dan Hrdy
 Yingzhi and **Xuedong**
(XD) Huang
 Nils and Yijing Sun
 Hultgren
 Scott J. Hultgren
Michael W. Hunkapiller
 George J. Isham
Irwin and Joan* Jacobs
 Anthony A. James and
 Cynthia K. French
 Kathleen Jamieson
Frank and Pam Joklik
 Fred Kavli*
Terri L. Kelly
 Kathryn S. and Peter S.
Kim
David B. Kirk and
 Caroline Landry
 Mary Anne Koda-Kimble

*Deceased

Gail F. Koshland
 Douglas E. Koshland
 Patricia Kuhl and Andrew
 Meltzoff
H.T. and Ling-Ling Kung
Wook Hyun Kwon and
 Son Cha Yi
 Lois and **Butler Lampson**
 Neal F. Lane
 Louis Lange
David C. Larbalestier
Edward D. Lazowska and
 Lyndsay C. Downs
Kai Li
 Jane and **Norman N. Li**
Frances and George
Ligler
 Christian Lütkemeyer
 James L. Manley
W. Allen Marr
 Kelsey C. Martin
Roger L. McCarthy
 Frank McCormick
 John F. McDonnell
Laura J. McGill
 Michael and Pat
 McGinnis
Robin K. and Rose M.
McGuire
Marcia K. McNutt
D. Nathan Meehan
Richard A. and Martha
 R. Meserve
David G. Messerschmitt

Maureen Miller*
 The Moore Family
Charles W. Moorman
G. Michael Morris
Cherry A. Murray
Jagdish Narayan
Ellen Ochoa
Thomas J. O'Neil
 Susan and **Franklin M.**
Orr, Jr.
 Thomas and Jo Overbye
 Larry* and Carol Papay
 David and Linda
Patterson
 Cathy Peercy
Fernando C. Pereira
 Jonathan B. and Donna
 J. Perlin
 Michael E. Phelps
Julia M. Phillips and
 John A. Connor
 Thomas and Patty Pollard
Dana A. Powers
 William H. Press and
 Jeffrey Howell
 Carol L. Prives
Stephen Quake
Richard F. and Terri W.
Rashid
Buddy Ratner and
 Cheryl Cromer
 Emanuel P. Rivers
Ronald L. Rivest
Anne K. Roby and

Lawrence Dodd
 Christopher Rogers
 Julie and **Alton D.**
Romig, Jr.
Howie Rosen and Susan
 Doherty
 Gerald M. Rubin
 William J. Rutter
Henry and Susan
Samueli
John M. Samuels, Jr.
Jerry Sanders III
Linda S. Sanford
Jose G. Santiesteban
Maxine L. Savitz
 Charles L. Sawyers
 Sophie Sawyers
 Mary Schafrik
John H. Schmertmann
 Wendy and **Eric Schmidt**
 Susan Schwartz
Charles L. Seitz
 Jacqueline and Henry
 Shinefield
Heung-Yeung Shum and
 Ka Yan Chan
Charles and Lisa Simonyi
Jennifer Sinclair Curtis
 Carol H. and Harold C.
 Sox
Robert F. and Lee S.
Sproull
Raymond S. Stata
 F. William Studier

Thomas and Marilyn
 Sutton
 Anne Swenson
Hemant K. and Suniti
Thapar
James M. Tien and Ellen
 S. Weston
William F. Tinney*
 Katherine K. and **John J.**
Tracy
Suzanne M. Vautrinot
Andrew and Erna*
Viterbi
John C. Wall
David Walt and Michele
 May
Darsh T. Wasan
 Irving L. Weissman
 Susan R. Wessler
 Ruth Williams and Bill
 Helton
Jie Xue
 Keith R. Yamamoto
Taiyin Yang
 Tilahun D. Yilma
 Anthony J. Yun and
 Kimberly A. Bazar
 Sandra Zaccaria
Elias A. Zerhouni
Charles F. Zukoski
 Anonymous (5)

Silver Level (\$5,000-\$9,999)

John E. Abele
Linda M. Abriola and
 Larry M. Albert
Ilesanmi and Patience
Adesida
 Vickie G. and **David J.**
Allstot
 Jeanne Altmann
Frances H. Arnold
Wanda M. and Wade
Austin
Ruzena K. Bajcsy

Terry and Roger N.
 Beachy
 Arthur L. Beaudet
Howard Bernstein
 Marie McCormick and
 Robert Blendon
Mark T. Bohr
 Jeanne Brooks-Gunn
 Michael S. Brown
John H. Bruning
 Robert Bryant and
 Réymundo Garcia

Martin J. Bukovac*
Christopher B. Burke
 Claude R. Canizares
 Christine K. Cassel
 Heide G. Castleman
Babu R. Chalamala
 Vicki L. Chandler
 Frank A. Chervenak
Margaret S.Y. Chu and
 Tze-Yao Chu
 Christos Coutifaris and
 Deborah Driscoll

Karen and **Steve Crouch**
 Johann Deisenhofer
 Carlos del Río and
 Jeannette Guarner
Hariklia Deligianni
 Sue K. Donaldson
 Gail D'Onofrio
 Robert D. Drennan
 Deborah Driscoll and
 Christos Coutifaris
 Anne M. Duderstadt
Abbas El Gamal

*Deceased

Janan and John Eppig
Bruce and Pat Finlayson
Stephen P. Fodor
Stephen R. Forrest
 Stephen P. Fortmann
 Elena and Peter
 Fuentes-Afflick
Douglas W. and Margaret
 P. Fuerstenau
Rajeev Gautam
 Julie L. Gerberding
Gary J. Goldberg
 Tony Gotto
Deborah L. Grubbe and
 James B. Porter, Jr.
 Mary Lou Guerinot and
 Rob McClung
Eliyahou Harari
 Arthur Hebard
Chris T. and Kathleen
 Hendrickson
 Katherine A. High
 Martha N. Hill
 Urs Höelzle
John R. Huff
 Tony Hunter
Mary Jane and Vernon
 Irwin
 Lisa P. Jackson
Michael R. Johnson
 Richard B. Johnston, Jr.
Robert E. Kahn and
 Patrice A. Lyons

Eric W. and Karen F.
 Kaler
Paul and Julie Kaminski
John and Wilma
 Kassakian
Farhad Khosravi
 Tadimitsu Kishimoto
 Richard D. Kolodner
 Edward A. and Kathryn
 F. Kravitz
Kent Kresa
Ellen J. Kullman
 Joseph Langsam
Enrique J. Lavernia
 Richard and Madeleine
 Lenski
 Alan and Agnes Leshner
 Marion Ein Lewin
 Peggy Mac Rae
Kiran Mazumdar-Shaw
Piotr D. Moncarz
 James A. Morone
 Jan and **E. Phillip***
 Muntz
Michael A. Mussallem
 Jaya and Venky
 Narayanamurti
Ronald and Joan
 Nordgren
Matt O'Donnell
 Gordon H. Orians
Roberto Padovani
David Parrillo

Wayne Pfeiffer
 Nancy N. Rabalais
 Srilatha and **Prabhakar**
 Raghavan
Marc H. Raibert
Elsa Reichmanis
John F. and Diane Reid
 Kenneth A. Ribet and
 Lisa R. Goldberg
 Ronald S. Robinson
 Judith Rodin
 Mary Ann and **Thomas**
 Romesser
Henry M. Rowan*
Barbara E. Rusinko
William and Emily
 Sanders
 Barbara Schaal and Joe
 Leverich
 Donna and **Jan Schilling**
Fred B. Schneider and
 Mimi Bussan
Julie M. Schoenung
Lyle H. Schwartz
Dominick M. Servedio
 Larry J. Shapiro
 Carla J. Shatz
 Michael Shelley
Deepika B. Singh
Larry L. Smarr
Arnold and Constance
 Stancell

David K. Stevenson
 Jerome F. Strauss III
Virginia and Carl
 Sulzberger
 Samuel O. Thier
 Linda Beth and James M.
 Tiedje
Jean W. Tom and
 Richard A. Register
 Katherine Tracy
John R. Treichler
 Susan Trumbore
David M. and Elaine Van
 Wie
Robert* and Robyn
 Wagoner
 Myrna M. Weissman
J. Turner Whitted
 Carl E. Wieman
 Charlton Wilson
Ward O. and Mary Jo
 Winer
 Michael S. Witherell
 Owen N. Witte
 Mary Woolley
Margaret M. Wu
Israel J. Wygnanski
Murty V.V.S. and
 Bhavani Yalla
Yannis and Sheryl Yortsos
Jie Zhang

Foundations, Corporations, and Other Organizations

In recognition of foundations, corporations, and other organizations that made gifts or grants of \$5,000+ to support the National Academy of Engineering in 2024.

American Endowment
 Foundation
 Aquanuity, Inc.
 Charities Aid Foundation
 of America
 Chevron Corporation
 Columbus Foundation
 Computers and
 Structures, Inc.
 CSX Corporation

Cummins, Inc.
 Golden Paws Foundation
 The Grainger Foundation
 Gratis Foundation
 Hoveida Family
 Foundation
 Hubbell Foundation
 Huff Family Foundation
 Jewish Community
 Foundation San Diego

Kelly Foundation
 Medtronic Foundation
 Minneapolis Foundation
 Mission Microwave
 Technologies
 The Gordon and Betty
 Moore Foundation
 Orange County
 Community
 Foundation

Philanthropic Ventures
 Foundation
 RobyDodd Charitable
 Foundation, Inc.
 Rogers Family Foundation
 Henry M. Rowan Family
 Foundation, Inc.
 Saint Louis Community
 Foundation
 Samueli Foundation

*Deceased

The San Francisco Foundation	Tien Family Foundation	Foundation, Inc.	Zerhouni Family
Tawny & Jerry Sanders Charitable Foundation	Tulsa Community Foundation	W.M. Keck Foundation	Charitable Foundation, Inc.
	United Engineering		

We have made every effort to list donors accurately and according to their wishes. If we have made an error, please accept our apologies and contact the Office of Development at 202.334.2431 or giving@nae.edu so we can correct our record.

Calendar of Meetings and Events

March 14	Hacking the Nervous System: The Convergence of Data, Devices, and Decision-making NAE Regional Meeting Brown University	May 6	NAE Council Meeting Virtual
		May 15	NAE Regional Meeting University of California, Berkeley
March 25–28	2025 German-American Frontiers of Engineering Oak Ridge National Laboratory	May 21	NAE Regional Meeting University of Minnesota and Medtronic, Minneapolis
April 2	Accelerating Clean Energy Manufacturing NAE Regional Meeting Georgia Institute of Technology	June 1–4	2025 Japan-America Frontiers of Engineering University of California, San Diego



LEADERSHIP IN A WORLD OF
Accelerating Change
CAMPAIGN FOR THE NATIONAL ACADEMY OF ENGINEERING

Use Your IRA to Support the NAE

Capitalize on a new, once-in-a-lifetime way to give through your Individual Retirement Account.

- You can establish a one-time \$54,000 Qualified Charitable Distribution-funded Charitable Gift Annuity.
- It will provide you with income.
- It may satisfy your Required Minimum Distribution.

Contact Elana Lippa, Director of Planned Giving, at 202.334.1817 or ELippa@nae.edu for information specific to you.

In Memoriam

Andreas Acrivos, 96, professor emeritus, Stanford University, died February 17, 2025. Dr. Acrivos was elected in 1977 for contributions in the application of mathematical analysis to the understanding of fundamental phenomena in chemical engineering processes.

John G. Anderson, 102, independent consultant, died November 22, 2024. Mr. Anderson was elected in 1979 for contributions in lightning research, extra and ultra-high voltage power transmission systems and power transformers.

Alfredo H.-S. Ang, 94, research professor of civil engineering, University of California, Irvine, died October 14, 2024. Dr. Ang was elected in 1976 for being a developer of practical and effective methods of risk and reliability approaches to engineering safety-and-design structural criteria formulation.

Donald L. Bitzer, 90, distinguished university research professor, North Carolina State University, died December 10, 2024. Dr. Bitzer was elected in 1974 for leadership in the utilization and development of technology for improving the effectiveness of education.

Kenneth E. Case, 80, regents professor emeritus, Oklahoma State University, died December 4, 2024. Dr. Case was elected in 1990 for significant contributions to statistical quality control through the incorporation of economic criteria in the design of quality systems.

Adrian R. Chamberlain, 94, independent consultant, died October 7, 2024. Dr. Chamberlain was elected in 2006 for innovations in the mobility, aesthetic, safety, and environmental aspects of transportation systems.

Joseph Feinstein, 88, retired consulting professor, Stanford University, died January 13, 2014. Dr. Feinstein was elected in 1976 for contributions to crossed-field electron devices.

Eli Fromm, 85, Roy A. Brothers University Professor Emeritus, Drexel University, died January 20, 2025. Dr. Fromm was elected in 2004 for innovation and leadership in the development of a holistic curriculum for engineering education.

Richard J. Grosh, 97, retired chairman and CEO, Ranco Management Corporation, died January 23, 2025. Dr. Grosh was elected in 1969 for significant contributions to heat transfer research.

Ronald A. Howard, 90, professor, Stanford University, died October 6, 2024. Dr. Howard was elected in 1999 for contributions to the foundations of decision analysis and its application.

Willi A. Kalender, 75, professor and chairman, University of Erlangen-Nuremberg, died October 20, 2024. Professor Kalender was elected an international member in 2016 for development of spiral computed tomography methods that enable modern high-speed 3-D medical imaging with X-rays.

Robert G. Loewy, 98, retired professor, School of Aerospace Engineering, Georgia Institute of Technology, died January 3, 2025. Dr. Loewy was elected in 1971 for contributions to the engineering of rotary-wing, vertical take-off and landing aircraft.

William R. Lucas, 102, retired director, NASA George C. Marshall Space Flight Center, died February 10, 2025. Dr. Lucas was elected in 1978 for contributions to research and engineering for space vehicles and the resolution of space flight problems.

Carl J. Monismith, 98, Robert Horonjeff Professor of Civil Engineering Emeritus, University of California, Berkeley, died January 7, 2025. Professor Monismith was elected in 1980 for contributions to asphalt paving technology which have led directly to improved methods of pavement design and rehabilitation.

Robin B. Nicholson, 90, Member of the Council, University of Exeter, died November 15, 2024. Sir Robin was elected an international member in 1983 for leadership in government/industry engineering policy, in bio-engineering, in theory of precipitation hardening in metals, and in solar energy materials systems.

Stuart B. Savage, 92, professor emeritus, McGill University, died November 26, 2024. Dr. Savage was elected in 2000 for contributions to the mechanics of granular flows that have laid the foundation for wide-ranging applications of particle technology.

Karl J. Springer, 89, retired vice president, Automotive Products and Emissions Research, Southwest Research Institute, died February 23, 2025. Mr. Springer was elected in 1996 for design of measurement and control systems to reduce smoke, odor, and other pollutants from diesel and gasoline engines.

George S. Springer, 90, Paul Pigott Professor of Engineering Emeritus,

Stanford University, died August 15, 2024. Dr. Springer was elected in 1994 for technical, educational and developmental contributions in composite materials for lightweight structures.

Charles V. Sternling, 92, retired research associate, Shell Oil, died November 1, 2017. Mr. Sternling was elected in 1976 for chemical engineering research on interphase

mass-transfer, and for solution of practical problems in two-phase flow and heat transfer.

Robert H. Wagoner, 73, professor emeritus, Ohio State University, died January 27, 2025. Dr. Wagoner was elected in 1983 for contributions to the materials mechanics and manufacturing technologies aspects of industrial metal forming.

Invisible Bridges

Ports, Politics, and Peace: The Engineering of Stability

Guru Madhavan is the Norman R. Augustine Senior Scholar and senior director of programs at the NAE.



Civilizations don't rise and fall by the sword alone. They are shaped—or shattered—by the infrastructure that anchors their networks. Infrastructure can foster peace and stability, but history shows it has also been a medium for influence. India's ancient ports made this clear: The sea wasn't just a bar-

rier but a conduit for empires to expand, where mutual exchange and leverage coexisted.

On Tamil Nadu's Coromandel Coast, Poompuhar (Kaveripoompattinam) flourished as a major port during the Sangam era (300 BCE to 300 CE). More than a trading hub, it was an engineered enclave of commerce, culture, and command, part of a maritime meshwork that linked South India to distant shores, weaving coastal economies into vast trade routes.

Under early Tamil dynasties and later the Cholas, Poompuhar thrived. Its layout reflected a hierarchy—fisherfolk and artisans worked by the shore, while merchants and rulers governed from inland quarters. Warehouses lined the waterfront, their goods marked with official seals, signaling a sophisticated regulatory system. A towering lighthouse beamed both welcome and watchfulness, ensuring commerce flourished—but on Chola terms. Tamil epics recount bustling ports where foreign merchants conducted business under local oversight, illustrating how infrastructure served as an economic enabler and an instrument of political control.

Nature, however, rewrote the script. As rising seas consumed Poompuhar in the early centuries CE, the Cholas turned their gaze southward. Over time, Nagapattinam's deep harbor positioned it as the region's dominant port, and by the medieval era, it had become a thriving nexus of commercial and cultural exchange. At its heart stood the Chudamani Vihara, a Buddhist monastery built in the 11th century with support from the Srivijaya empire of present-day Indonesia. Its soaring spires welcomed mariners from across the Bay, while its halls hosted monks from Sri Lanka, China, and Southeast Asia, reinforcing Nagapattinam's role as both a crossroads and a conduit of Tamil reach—which flowed not just through fleets but through ideas and institutions. The Chola-Srivijaya partnership wasn't merely symbolic—it deepened regional stability through shared infrastructure and interconnected trade routes.

Today, infrastructure shapes global leverage, forging interdependence in ways both subtle and stark. The Greater Mekong subregion's power grid synchronizes economies, yet its benefits remain unevenly distributed. Once divided by water disputes, Thailand and Laos now exchange electricity, intertwining their energy needs and reinforcing cooperation through both politics and engineering. Similarly, the Trans-Asian Railway stitches together China, Laos, and Europe, turning old frontiers into transit corridors. But mismatched rail gauges and aging networks still pose hurdles, proving that infrastructure is as much about negotiation as construction.

Water itself has become a medium of diplomacy. The Indus Waters Treaty shows how engineering details can turn potential competing interests into cooperation. When Pakistan challenged India's Kishanganga hydroelectric project in 2010, the resolution didn't come through political posturing but through numerical negotiations—flow rates, reservoir capacities, and power generation schedules. Today, hydroelectric projects threading the Chenab and Jhelum rivers continue this approach, demonstrating that stability is often engineered, not just brokered. Such agreements depend not only on political will but on the technical infrastructure that makes cooperation possible—dams, grids, servers, and satellites.

In the South China Sea, key ports reinforce sovereign

⁵ This piece was previously published by the Asian Peace Programme of the Asia Research Institute at the National University of Singapore and in *The Straits Times* (Singapore) on March 6, 2025. Thanks to Ambassador Kishore Mahbubani for inspiring and inviting this perspective.

claims. Radar stations track naval movements. Deep-draft harbors accommodate warships and cargo vessels alike. Despite territorial disputes, commerce flows through shared shipping routes, standardized navigation systems, and cooperative security measures. As with Poompuhar, infrastructure anchors both strategic presence and partnership.

China's Belt and Road Initiative (BRI) takes this further. While some projects in Pakistan have faced setbacks, the Jakarta-Bandung rail and Laos-China railway demonstrate successful partnerships. Beyond construction, these projects entrench long-term economic ties through maintenance agreements, operational protocols, and financial dependencies. Through the BRI, China transforms trade routes into economic strongholds during a pivotal period of regional competition. Similarly, as Arctic ice recedes, the Northern Sea Route and Transpolar Sea Route are redrawing access, with Russia expanding northern ports and deploying nuclear icebreakers to establish year-round control. Just as shifting coastlines altered Poompuhar's fate, these changing pathways and corridors determine not just the patterns of collaboration, but the distribution of regional leverage.

Infrastructure dictates far more than logistics in Southeast Asia. The Johor-Singapore Causeway, for instance, does more than move traffic—emergency response teams share protocols, logistics systems operate in sync, and border controls function as one. Singapore's Tuas Port embodies this precision, an automated marvel built for speed and scale. AI-powered cranes hoist containers with precision, autonomous vehicles zip across the docks, and intelligent berthing systems orchestrate seamless arrivals, handling 10 million shipping containers, or twenty-foot equivalent units (TEUs), since the start of operations in 2022, with new capabilities to handle 65 million TEUs when fully operational in the 2040s, making it the world's largest automated port. Every transnational project—whether a railway in Africa, an airport in Asia, or a data corridor in the Americas—rescripts influence without a single shot fired.

Like South India's ancient ports, these choices outlast rulers and regimes, binding nations through networks of mutual reliance. Peace and geopolitical stability aren't just diplomatic achievements; they are engineered, and infrastructure is often the stabilizer that endures—even outlives—shifting alliances.

The BRIDGE

(USPS 551-240)

National Academy of Engineering
2101 Constitution Avenue NW
Washington, DC 20418

Periodicals

Postage

Paid

**NATIONAL
ACADEMIES** *Sciences
Engineering
Medicine*

The National Academies provide
independent, trustworthy advice
that advances solutions to society's
most complex challenges.

www.nationalacademies.org