Stanford University
Human-Centered
Artificial Intelligence

Stanford
University

Annual Report 2024

Stanford Institute for Human-Centered Artificial Intelligence

In its fifth year, HAI continued to accelerate its efforts to bring together thought leaders from academia, industry, government, and civil society to shape the development and responsible deployment of AI.

Contents

Letter from the HAI Leadership Team

Five years ago, we began a journey with a bold vision: to guide and build AI that improves the human condition. HAI was the first public sector organization established to drive innovation and engage policy, industry, and civil society around the mission of human-centered AI. Today, HAI is firmly established as a trailblazer in this space, shaping the ethical and inclusive development of AI technologies. We are thrilled to have exceeded expectations, driving real-world impact through our many research, education, and policy programs.

Among our five-year achievements, we are especially proud to share these highlights:

- Since 2019, HAI has awarded $45 million in funding to faculty across all of Stanford's seven schools, in fields ranging from robotics and cognitive science to healthcare and education. Of that, approximately $12 million in Seed Grant awards has attracted over $25 million in external funding, creating a multiplier effect for many of our affiliated scholars.
- Forecasting a need to study the economy and the emerging foundation model space, HAI established two specialized research centers: the Stanford Digital Economy Lab (DEL), which explores AI's impact on the economy and nature of work; and the Center for Research on Foundation Models (CRFM), which seeks to understand and shape the development of foundation models.
- With the seventh edition published in 2024, the annual AI Index Report has become the definitive resource for AI macro-trends and analysis, while the Foundation Model Transparency Index, introduced in October 2023, exposed an alarming lack of transparency among the top 10 foundation model developers.
- HAI envisioned and championed the CREATE AI Act (Creating Resources for Every American to Experiment with Artificial Intelligence Act) to establish a National AI Research Resource (NAIRR) that proposes to make costly compute resources and government datasets widely available to academics, nonprofit researchers, and start-ups so they can conduct education and research for the public good. Through this process, HAI has shaped global conversations about the need to provide resources to more colleges, universities, and nonprofits in order to ensure more equitable access and opportunities for academia and the public sector to shape the future of human-centered AI.
- In the 2023–24 academic year, HAI made notable progress in our policy engagement and thought leadership efforts. We tailored our education programs to meet the needs of distinct audiences, including undergraduates, executives, government officials, and social sector employees. We increased engagement with the broader AI community through our content and media relations efforts.

Looking ahead, we can see that in many ways, AI is pushing the research boundaries of the university model. To ensure the next generation of this technology gets created for the public good, academic scholars will need access to deeper resources. This is why HAI is advocating for a collaborative lab environment funded by philanthropy and government support. We envision a setting in which students have access to cutting-edge training and the possibility to go deep into researching complex AI models for the benefit of humanity.

As we continue this exciting journey, we are grateful for the support and partnership of our extended HAI community. We look forward to building a future together that reflects our highest potential as well as our shared values as a society.

Fei-Fei Li

Sequoia Professor, Computer Science Department

Denning Co-Director, Stanford HAI

James Landay

Anand Rajaraman and Venky Harinarayan Professor, School of Engineering, Professor of Computer Science

Denning Co-Director and Faculty Director of Research, Stanford HAI

John Etchemendy

Patrick Suppes Family Professor in the School of Humanities and Sciences

Denning Co-Director, Stanford HAI

Russell Wald

Executive Director, Stanford HAI

This annual report reflects the research, activities, and impact of Stanford HAI between September 2023 and August 2024.

Spotlight on HAI at Five: Celebrating 5 Years of Impact

In the five years since our founding in 2019, HAI has fostered an interdisciplinary community of people in research, industry, policy, and civil society who have helped to carry forth our vision for promoting human-centered uses of AI. In spring 2024, we invited this community to join us in celebrating HAI's fifth anniversary with a conference focused on the theme, "HAI at Five: Celebrating 5 Years of Impact."

Throughout the day, thought leaders and experts in the field covered a wide range of issues, including mapping the human brain, removing impediments to a robust AI innovative ecosystem, accelerating science with AI, and implementing AI in society at scale. A virtual poster session showcased the work of the HAI community over the last five years, highlighting more than 20 projects that HAI has supported to date.

CREDIT: CHRISTINE BAKER

HAI Timeline

2019

Mar: HAI officially launches
Mar: HAI proposes a national AI research resource to drive American innovation in AI
Apr: HAI launches fellowship program for faculty

2020

Feb: Economist Erik Brynjolfsson joins HAI, launches the Stanford Digital Economy Lab (S-DEL)
Aug: Hoffman Yee Grants launch
Oct: HAI, with Stanford CS and the Center for Ethics in Society, launches Embedded EthiCS

2021

Jun: HAI pilots Ethics & Society Review for awarding grants
Aug: HAI's Center for Research on Foundation Models launches
Aug: HAI creates graduate fellowship program
Oct: HAI and Stanford scholars develop a blueprint for a NAIRR

2022

Aug: HAI launches policy boot camp
Sep: HAI launches AI startup series
Sep: HAI debuts student affinity groups
Dec: HAI starts Corporate Affiliate Program

2023

Jan: HAI partners with Wu Tsai Center to fund research in AI and neuroscience
Mar: HAI develops Tech Ethics & Policy fellowships
Jun: Stanford Medicine and HAI announce RAISE-Health
Jun: HAI leaders Fei-Fei Li and Rob Reich meet with President Biden
Sep: HAI launches the AI Training Series for Federal Employees
Nov: HAI and The Asia Foundation announce AI Perspectives from Asia

HAI Mission & Research Focus

Stanford HAI's mission is to advance AI research, education, policy, and practice to improve the human condition. We believe AI should be guided by its human impact, inspired by human intelligence, and designed to augment, not replace, people. Each of our research initiatives connects to one of these three pillars:

CREDIT: CHRISTINE BAKER

Human Impact

Human Impact projects seek to develop a deeper understanding of how we can address the issues society will confront as AI becomes commonplace.

Augment Human Capabilities

Research efforts that augment human capabilities have the potential to educate, train, and support individuals in ways that lead to progress in health care, education, sustainability, automation, and other domains.

CREDIT: CHRISTINE BAKER

Intelligence

Intelligence initiatives aim to develop AI that understands human language, emotions, inventions, behaviors, and interactions at multiple scales.

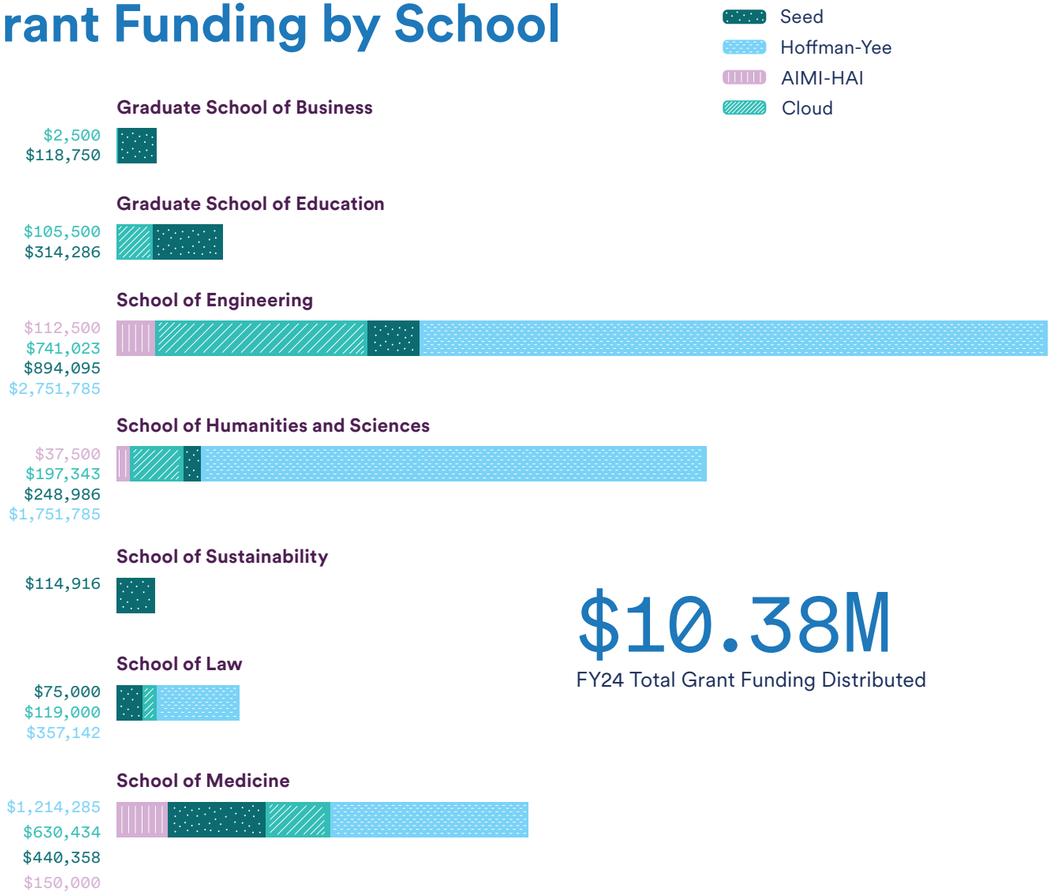
Research

HAI's research team supports its mission through funding innovative and multidisciplinary research, convening scholar workshops, and producing the annual AI Index.

Grant Programs for AI Research

In AY 2023–24, HAI distributed $10.3 million to faculty spanning all seven Stanford University Schools to work on projects aligned with the three research focus areas of human impact, augment human capabilities, and intelligence. Many of these projects span multiple departments in keeping with the institute's commitment to support interdisciplinary AI research. Each year, this funding generates a multiplier effect. At publication time, HAI grant recipients reported $10.8 million in follow-on funding in AY 2023–24.

Grant Funding by School



$10.38M
FY24 Total Grant Funding Distributed

"From my perspective, the HAI funding programs have been excellent and have had a positive impact on research and collaboration among affiliated faculty."

Steve Collins
Mechanical Engineering

"The funding fosters interaction across campus, which often leads to novel insights and viewpoints on a problem."

Jeannette Bohg
Computer Science

HAI runs three types of grant programs with specific and disparate objectives:

01. Hoffman-Yee Grants

The Hoffman-Yee Grants, generously supported by Reid Hoffman and Michelle Yee, fund "asteroid-shot" ideas to address significant scientific, technical, or societal challenges. HAI believes the results of these projects will play a significant role in defining future work in AI, from academia to industry, government, and civil society. This year's cohort of six interdisciplinary teams received $500,000 each to pursue their research projects, and four of the teams were selected for follow-on funding of $1 million to $2 million over the next two years.

MEDIA MENTIONS

AI's openness is being sharply debated by technologists, policymakers
SF Examiner, June 16, 2024

What Gen Z Wants from AI Policymakers
Semafor Tech, June 7, 2024

US National Security Experts Warn AI Giants Aren't Doing Enough to Protect Their Secrets
Wired, June 6, 2024

To solve AI's energy crisis, 'rethink the entire stack from electrons to algorithms,' says Stanford prof
The Register, June 5, 2024

Meet 12 Asteroid Shots in AI
Machine Learning, December 11, 2023

EVENT HIGHLIGHT

In September 2023, HAI hosted the Hoffman-Yee Research Symposium, which showcased six winning teams' progress toward realizing their ambitious ideas. With more than 200 people joining in person and over 250 virtual participants, the event showcased advances in areas such as:

- Integrating technically better, socially responsible foundation models
- Bridging the gap between vision and language models
- Building the next-generation, brain-inspired computer chip
- Improving immigration integration
- Encoding societal values expressed on social media
- Developing a framework for children's diabetes care

CREDIT: DAVID GONZALES

CREDIT: DAVID GONZALES

PROJECT HIGHLIGHT

Tuning Our Algorithmic Amplifiers: Encoding Societal Values into Social Media Algorithms

Michael Bernstein, Main PI
School of Engineering, Computer Science

Nathaniel Persily, Co-PI
School of Law, Law School

Angele Christin, Co-PI
School of Humanities and Sciences, Communication

Jeanne Tsai, Co-PI
School of Humanities and Sciences, Psychology

Jeffrey Hancock, Co-PI
School of Humanities and Sciences, Communication

Johan Ugander, Co-PI
School of Engineering, Management Science and Engineering

Tatsunori Hashimoto, Co-PI
School of Engineering, Computer Science

This team is seeking to develop intertwined social scientific, engineering, and policy answers to the question, "How can we encode societal values into these algorithms without sacrificing the core of what can make social media compelling?"

The researchers have created a first-in-class social media feed algorithm that translates social science research on anti-democratic attitudes into a concrete formalization and minimizes it as part of a news feed-ranking algorithm. In an online experiment, the team empirically demonstrated that this algorithm reduced political animosity among nearly 2,000 U.S. partisans without decreasing engagement or prompting perceived threats to freedom of speech. The team's findings are described in an April 2024 paper that was accepted into the ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW).

PROJECT HIGHLIGHT

EAE Scores: A Framework for Explainable, Actionable, and Equitable Risk Scores for Healthcare Decision

Carlos Ernesto Guestrin, Main PI
School of Engineering, Computer Science

Carissa Carter, Co-PI
School of Engineering, d.school

Emily Fox, Co-PI
School of Humanities and Sciences, Statistics, Computer Science (courtesy)

Ramesh Johari, Co-PI
School of Engineering, Management Science and Engineering, Electrical Engineering (courtesy), Computer Science (courtesy)

David Maahs, Co-PI
School of Medicine, Pediatrics

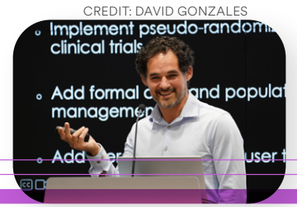
Priya Prahalad, Co-PI
School of Medicine, Pediatrics

Sherri Rose, Co-PI
School of Medicine, Health Policy

David Scheinker, Co-PI
School of Medicine, Pediatrics

CREDIT: DAVID GONZALES

CREDIT: DAVID GONZALES



This team is building a "Framework for Explainable, Actionable, and Equitable Risk Scores for Healthcare Decisions" with a focus on diabetes care.

The scholars have made significant progress toward three goals: sensor-enabled monitoring of patient state; algorithmic tools and technology for optimizing the targeting of care resources and interventions depending on patient state; and AI methods that work in partnership with human clinicians, rather than supplanting them, to enable remote patient monitoring at scale. A key outcome for this team over the past year was the publication of "Smart Start: Designing Powerful Clinical Trials Using Pilot Study Data," a three-step method for using retrospective data from a pilot study to design an adaptive health intervention and a subsequent clinical trial, in the New England Journal of Medicine.

02. Seed Research Grants

The HAI Seed Research Grants, supported early on by Steve and Roberta Denning and for the past four years by Dalio Philanthropies, are dedicated to supporting new, ambitious, and speculative ideas with the objective of getting initial results. Since the program's start in 2018, approximately $12 million in awards has attracted over $25 million in external funding. This outcome demonstrates the program's promise in identifying high-reward, speculative, and early-stage research. Additionally, due to the breadth of proposals the institute receives from across campus, the seed grants provide a bellwether for what is to come in the field of AI.

PROJECT HIGHLIGHT

Force-based Robot Learning for Soft Tissue Manipulation

Allison Okamura, Main PI
School of Engineering, Mechanical Engineering

Carla Pugh, Co-PI
School of Medicine, Surgery

Jeannette Bohg, Co-PI
School of Engineering, Computer Science

This project explored incorporating an awareness of force into robot-assisted surgery to augment the skills of human surgeons. The team theorized that a robot that can gently manipulate tissue during surgery could be incredibly valuable. The researchers trained two robotic arms: one using vision and motion data, and the other using vision, motion, and force data. The arm trained with force data was three times more successful and applied significantly less force to the tissue.

The results were presented at a leading medical robotics conference, and the team has submitted a proposal for an Exploratory/Developmental Grant (R21) from the National Institutes of Health. The team received a donation of a new da Vinci Research Kit from Intuitive Surgical Inc., which will be housed in the Stanford Robotics Center.

03. Cloud Credit Grants

Cloud Credit Grants support emerging research that requires advanced computational resources provided by the commercial cloud. In the last academic year, HAI distributed $1.8 million in cloud credits with generous gifts from Microsoft Azure and Google Cloud.

PROJECT HIGHLIGHT

Controllable Language Modeling with Discrete Diffusion

With compute resources from Google Cloud Credits, Stefano Ermon (Computer Science) and colleagues developed a novel approach to training discrete diffusion models called Score Entropy Discrete Diffusion. When the team tested the method on standard language modeling tasks, it outperformed GPT-2. This work won best paper at the International Conference on Machine Learning (ICML) 2024.

Ethics & Society Review

Before receiving funding, all research proposals submitted to HAI undergo multiple rounds of rigorous faculty review, including an Ethics and Society Review (ESR). The ESR assesses potential negative impact on society before being greenlighted for funding.

ESR conducted a mixed-methods evaluation, described in a Proceedings of the National Academy of Sciences (PNAS) paper, in which surveys and interviews of researchers found:

100%
were willing to continue submitting future projects to the ESR

58%
felt that it had influenced the design of their research project

The ESR process has been so successful that it is now being used in other research domains at Stanford, beyond HAI and AI.

> "Fast review to enable accelerated research on new topics, overall process was excellent, the addition of ethical review is very important."
>
> **Allison Okamura**
> Mechanical Engineering

State of AI Reports

AI100

The One Hundred Year Study of Artificial Intelligence (AI100) is a 100-year effort to study and anticipate how AI will affect every part of human life. The study is administered by HAI and managed by a standing committee of AI leaders from institutions around the world, chaired by Vincent Conitzer at Carnegie Mellon University. AI100 releases reports every five years with the goal of documenting AI developments and their ethical implications for society, as well as providing a through-line that marks the trajectory of the field.

During the past academic year, AI100 selected Mike Wooldridge, Professor of Computer Science, University of Oxford, to lead the Study Panel tasked with writing the next report. The Standing Committee also added three new members: David Autor, Professor of Economics, Massachusetts Institute of Technology; Ryan Calo, Professor of Law, University of Washington; and Yejin Choi, Professor of Computer Science, formerly at the University of Washington, joining Stanford University in 2025.

AI Index

HAI and the AI Index Steering Committee released the 2024 AI Index report in April 2024. This seventh edition featured data from a broad range of academic, private, and nonprofit organizations, along with more self-collected data and original analysis than any previous report.

This year's launch included several high-profile events, such as a panel discussion at the Council on Foreign Relations, a briefing for U.S. congressional staffers, and a workshop for the Swedish AI Commission. The 2024 AI Index website has been viewed over 600,000 times, and the report has been downloaded 135,000 times. Additionally, since its inception, the AI Index has been referenced by governments around the world, including those of the United States, China, and the United Kingdom.

Since Launch:

500+
pieces of news coverage

600,000+
website page views

50+
promotional events to industry, government, & academic audiences

135,000+
downloads

MEDIA MENTIONS

A.I. Has a Measurement Problem
The New York Times, April 15, 2024

An auspicious 'World Quantum Day'
Politico, April 15, 2024

Is there enough text to feed the AI beast?
Semafor, April 16, 2024

AI-Related Regulations Rise in US and Globally: Stanford Report
Bloomberg Law, April 15, 2024

Responsible AI's yardstick mess
Axios, April 16, 2024

Stanford HAI on 2024 AI Global Index Report
Bloomberg, April 16, 2024

Mind and Iron: AI is hitting a wall its champions are in denial about
Mind and Iron, April 18, 2024

Are We Running Out of Training Data?
The Information, April 20, 2024

Generative AI is speeding up human-like robot development. What that means for jobs
CNBC, May 7, 2024

Research Convenings

AI + Health 2023 Online Conference

HAI co-organized the third annual AI + Health online conference in collaboration with the Center for Artificial Intelligence in Medicine and Imaging (AIMI) and the Center for Continuing Medical Education (CME). The December 2023 event convened 1,000 experts and leaders from academia, industry, government, and clinical practice to explore critical and emerging issues related to AI's impact across the spectrum of healthcare.

Three conference tracks included:

- Advancing Clinical Excellence and Healthcare Delivery
- Advancing Health AI Research & Development
- New AI Methods to Accelerate Innovation in Healthcare

AI + Education Summit: Advancing Human Learning with AI Technologies

Building on the success of the first summit in 2023, HAI and the Stanford Accelerator for Learning hosted the second summit in February 2024. The event brought together Stanford faculty and students, along with leaders and participants from academia, industry, Civil society, PK–12 education, and government. As AI technologies continue to evolve and new opportunities and challenges emerge in education, the AI + Education research community showcased the latest findings and solutions to meet key challenges.

CREDITS: JEANINA MATIAS

Education

The mission of HAI's education portfolio is to equip leaders with AI knowledge, helping them address the societal and ethical implications of AI across government, business, and education. In academic year 2023–24, HAI developed and continued the following programs to strengthen our community of young scholars as well as experienced professionals.

Student Affinity Groups

This year marked the second cohort of the student affinity group program; 12 teams of interdisciplinary students participated (70+ students). Each group formed around a distinct interest in a topic related to the development or study of human-centered AI. Teams received up to $1,000 in funding for the academic year.

EVENT HIGHLIGHT

Lightning Talk Session

The groups shared their outcomes during a lightning talk session in spring 2024. For example, one of the teams examined the ethical challenges faced by data workers and the companies that employ them.

CREDIT: JEANINA MATIAS

Graduate Fellows

HAI offers a three-quarter fellowship program for Stanford graduate students to encourage interdisciplinary research conversations, facilitate new collaborations, and grow the HAI community of graduate scholars who are working in areas related to AI. The program fosters collaboration among engineers, social scientists, humanists, and others researching the future of purposeful, intentional, and human-centered AI.

In AY 2023–24, HAI hosted 19 graduate students from six Stanford schools. These fellows participated in programming that emphasized the human-centered AI perspective from HAI's faculty leaders through biweekly seminars and a capstone project.

AI Literacy in K–12 Schools

HAI's K–12 programs focus on enhancing AI literacy for educators and students from diverse backgrounds. For example, the AI4ALL summer camp immerses underrepresented students in hands-on AI research and mentoring at Stanford to increase diversity in the field's future leaders. Together, these K–12 initiatives equip teachers and youth with the knowledge, skills, and access needed to shape AI innovations in ways that reflect community needs.

Human-Centered AI Courses

HAI faculty and fellows bring diverse perspectives on AI technology and its societal impact to Stanford students and adult learners through innovative courses and programs.

CS 139: Human-Centered AI

Peter Norvig
HAI Distinguished Education Fellow

In its third year, *CS139: Human-Centered AI*, led by **Peter Norvig** and Dan Russell, emphasized the intersection of AI development and human impact, delving into how advanced tools must be designed with consideration for their effects on users and communities. The course was recorded and is intended to become a signature HAI online course for professionals.

ENGLISH 106A: Black Mirror: A.I. Activism

Michele Elam
HAI Senior Fellow

This small-group course led by **Michele Elam** explores the social, ethical, and artistic implications of AI systems with an emphasis on aesthetics, civic society, and racial justice, including scholarship on decolonial AI, indigenous AI, disability activism AI, feminist AI, and the future of work for creative industries.

MUSIC 356: Music and AI (CS 470)

Ge Wang
HAI Senior Fellow and Associate Director

In this "critical making" course led by **Ge Wang**, students learn practical tools and techniques for AI-mediated music creation, engineer software systems incorporating AI, HCI, and music, and critically reflect on the aesthetic and ethical dimensions of technology.

Professional Education

Offered in person and virtually, HAI's professional education programs cultivate synergy and strengthen connections between academic endeavors and industry in alignment with HAI's mission. These programs provide a platform for industry leaders to engage with real-world use cases. In 2023–24, HAI delivered executive education programs to a diverse set of organizations, including Accenture, PepsiCo, and Ernst & Young.

Responsible AI Program for Accenture

HAI conducted this on-campus program over three consecutive days. Attendees included 65 international Accenture executives. During the program, attendees cultivated a shared understanding of responsible AI, informed by insights from industry and academia. This course empowered participants to develop operational expertise and helped them build confidence for navigating the ethical complexities of AI.

Generative AI: Technology, Business, and Society Program

This program, developed with Stanford Online, explores the technical fundamentals, business implications, and societal considerations of AI with a focus on putting people first.

Policy

In AY 2023–24, HAI further solidified its role as a global leader in shaping AI governance. HAI's policy team drove key conversations around the responsible development, application, and governance of AI. Through engagements with domestic and international policymakers, HAI shared its cutting-edge, multidisciplinary empirical research, tailored educational programs, and convenings with diverse stakeholders. HAI policy efforts are made possible by philanthropic support from entities such as the Patrick J. McGovern Foundation and donors like Bob and Dottie King.

Policy Research

Through papers and briefs, HAI draws multidisciplinary AI policy insights from Stanford's cutting-edge research. These translational and analytic pieces equip policymakers with tools to better understand and govern AI, ensuring policies are informed by the latest research and aligned with real-world needs.

To stay updated on the latest insights, research, and opportunities on the intersection of policy, society, and AI sign up for the HAI policy newsletter.

Policy Briefs

Policy briefs translate the key findings from published academic articles on policy-relevant AI topics into succinct recommendations for regulators and other audiences. HAI's 12 published policy briefs in the last academic year were widely received. For example, Understanding Liability Risk from Healthcare AI, coauthored by HAI faculty affiliate Michelle Mello, coincided with her testimony before the U.S. Senate, enabling immediate congressional engagement on healthcare AI regulation. Other briefs on topics such as The AI Regulatory Alignment Problem and Foundation Models and Copyright Questions also garnered significant attention.

CREDIT: PEPE GOMEZ

Responses to Requests for Information

In response to federal requests for information or comment, HAI scholars provided policymakers with recommendations to consider when drafting rulemaking or federal agency initiatives. Those included Managing Misuse Risk for Dual-Use Foundation Models, Response to USAID's Request for Comment on AI in Global Development Playbook, and Responses to OMB's Request for Comment on Draft Policy Guidance on Agency Use of AI.

Original Policy Research

HAI policy white papers present in-depth research and bold proposals on critical AI policy issues to policy audiences.

HAI's white paper, Exploring the Impact of AI on Black Americans: Considerations for the Congressional Black Caucus's Policy Initiatives, prepared for the Congressional Black Caucus in conjunction with Black in AI, equips lawmakers with the tools to develop a thoughtful, forward-looking AI policy strategy that ensures the benefits of this technology are widely shared and its risks are carefully managed. Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World explores the current and future impact of privacy and data protection legislation on AI development and provides recommendations for mitigating privacy harms in an AI era.

HAI scholars have analyzed AI-related U.S. executive actions issued since 2019. Following Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of AI, HAI, CRFM, and RegLab developed a detailed tracker of its 150 implementation requirements. This resource helps policymakers and the public monitor progress and understand federal implementation efforts.

Research Workshops

In closed-door policy research workshops, HAI convenes leading policymakers, academics, practitioners, and members of civil society for candid and outcome-focused multi-stakeholder discussions on critical AI policy issues. The workshops lay the foundation for more in-depth policy research and tailored policy recommendations.

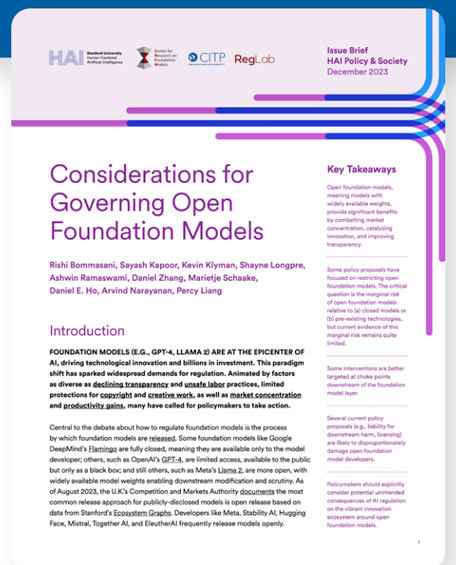
Healthcare AI Policy Workshop HAI organized a workshop with 50 leading policymakers, scientists, healthcare providers, ethicists, AI developers, and patient advocates to discuss regulatory challenges in healthcare AI across three use cases. The workshop was led by HAI Associate Director Curt Langlotz with HAI's Healthcare AI Policy Steering Committee.

Global South AI Workshop HAI held a two-part workshop with 32 civil society and tech leaders to discuss AI's impact on elections in South Asia, Latin America, and Africa—and to explore ways to strengthen the AI ecosystem in the Global South. The workshop was co-hosted with Stanford's Cyber Policy Center and The Asia Foundation.

RESEARCH HIGHLIGHT

Considerations for Governing Open Foundation Models

In response to the growing policy discussions on open-source AI, HAI, in collaboration with CRFM, RegLab, and Princeton's Center for Information Technology Policy, developed an issue brief addressing the benefits and risks of open foundation models and introducing a framework to assess their marginal risks. To further this dialogue, HAI convened a workshop with 42 participants from the federal government, industry, and civil society. Policymakers, including the U.S. National Telecommunications and Information Administration, have widely adopted this framework. This work culminated in a publication in Science, advancing a more nuanced understanding of open foundation models and their policy implications.


Issue Brief
HAI Policy & Society
December 2023

Considerations for Governing Open Foundation Models

Rishi Bommasani, Sayash Kapoor, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaake, Daniel E. Ho, Arvind Narayanan, Percy Liang

Introduction

Key Takeaways

Tech Ethics & Policy Summer Fellowships

For the second year, HAI awarded 12 Tech Ethics & Policy fellowships to Stanford graduate students in technical programs as part of our efforts to build a stronger technical talent pipeline into government and public service. The program demonstrates to early-career technologists how valuable their diverse STEM skillsets are to informing public policy and sound AI governance. Through a specialized spring course and a summer placement in Washington, D.C., fellows gained firsthand experience working in congressional offices, government agencies, and think tanks. The student fellows were also connected to Stanford alumni mentors working in Washington D.C. to provide guidance and support. The program, run in collaboration with the McCoy Family Center for Ethics in Society, is supported by a gift from Project Liberty.

For the 2024 cohort, placements included:

- American Enterprise Institute
- Center for Democracy and Technology
- Department of Commerce
- Department of Health and Human Services
- Freedom House
- National Oceanic and Atmospheric Administration
- Senate Committee on Homeland Security & Governmental Affairs
- U.S. AI Safety Institute
- The Brookings Institution
- White House Office of Management and Budget

Policy Engagement

HAI significantly increased its policy engagement this year to meet growing demand from policymakers seeking expert insights on AI innovation and governance. The team deepened its collaboration with policymakers in California, the United States, European Union, United Nations, and other international organizations and global governments.

HAI experts testified on issues of AI governance before multiple federal and state legislative committees, including the U.S. Senate Committee on Homeland Security and Governmental Affairs, the U.S. Senate Committee on Finance, the U.S. House Oversight Subcommittee on Cybersecurity, Information Technology, and Government Innovation, a joint hearing of two California State Legislature committees, the California State Senate Governmental Organization Committee, the California State Assembly Privacy and Consumer Protection Committee, and the California Little Hoover Commission. HAI also engaged with a variety of policymakers through meetings, workshops, and other forums noted below:

By the numbers:

18
U.S. federal agencies

13
California state agencies and legislative offices

42
U.S. Congressional offices and committees

15
Foreign government offices

Policy Convenings

HAI's policy convenings offer a vital forum to unite leaders from academia, industry, government, and civil society, creating opportunities for deep discussions and collaboration across disciplines. By bringing together diverse perspectives, HAI facilitates meaningful dialogue on shaping the future of AI governance and its societal implications.

AI+Policy Symposium: A Global Stocktaking

This symposium was hosted by HAI and the Stanford Cyber Policy Center and led by Marietje Schake and James Landay, with 230 in-person participants for a day-long gathering to discuss the state of global AI policymaking. Participants included officials and policy leaders from the United States, United Kingdom, European Union, Brazil, China, India, South Africa, Organization for Economic Co-operation and Development (OECD), African Union, and more.

CREDIT: PATRICK BEAUDOUIN

Joint California Summit on Generative AI

HAI, along with co-organizers—UC Berkeley College of Computing, Data Science and Society, and the California state government—hosted this summit in May 2024. Stanford faculty Dan Ho, Fei-Fei Li, and Chris Manning, plus incoming university president Jonathan Levin, participated in panels, including a fireside chat featuring Li and Gov. Gavin Newsom.

Education & Training for Policymakers

HAI provides educational opportunities to close the gap between rapid AI advancements and policymakers' understanding of the technology. Through tailored training, these programs equip leaders with knowledge and tools to make informed decisions and develop responsible AI policies.

2024 Congressional Boot Camp on AI

HAI launched its congressional boot camp in 2022 to fill a perceived gap in exposure to cutting-edge, technical research on developments in the AI field. In three years, the program has enabled over 75 congressional staffers from both sides of the aisle to engage directly with HAI faculty researchers to learn about the policy implications of work from a range of interdisciplinary backgrounds. HAI's 2024 boot camp brought 26 bipartisan congressional staffers to Stanford for three days, with sessions on AI's impact on international security, work, and healthcare, featuring scholars, Silicon Valley leaders, and civil society pioneers.

CREDIT: DANIEL BECK

AI Training Series for Federal Employees

In October 2023, HAI partnered with the U.S. General Services Administration and the White House Office of Management and Budget to develop a specialized training program for government employees at the local, state, and federal level. This six-session program introduced key AI concepts—including technical foundations, public sector use cases, and risk mitigation strategies—to over 8,000 federal employees who registered for the course. The series supports the mandate from the AI Training Act of 2022 and empowers participants with the knowledge to navigate and manage AI in their work. This was HAI's second year participating in the training.

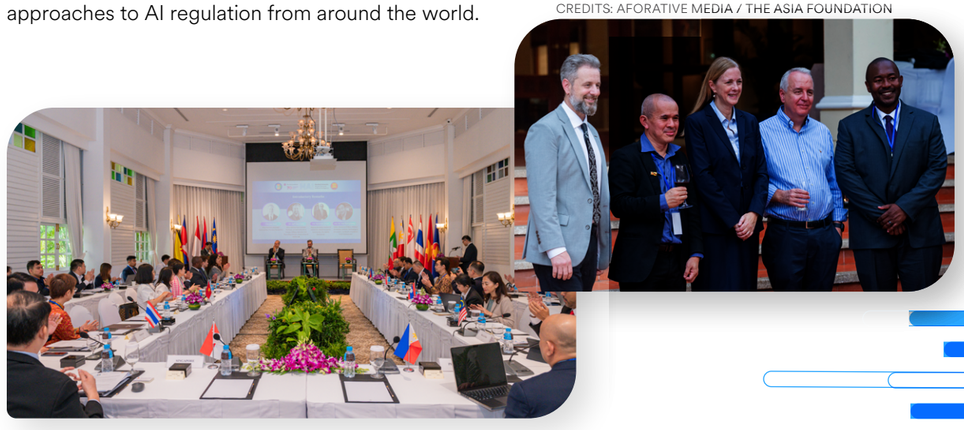
Workshop with the Association of Southeast Asian Nations (ASEAN)

In June 2024, HAI collaborated with The Asia Foundation to host digital ministry officials, technologists, and members of the ASEAN community for a workshop on the eve of ASEAN's Committee on Science, Technology & Innovation Ministerial Meeting on Science, Technology, and Innovation in Siem Reap, Cambodia. HAI faculty affiliates Jeff Hancock and Sanmi Koyejo discussed AI governance, bias, fairness, and trust, and they convened a panel of policy experts to discuss approaches to AI regulation from around the world.

AI Fundamentals for Public Servants

In August 2024, HAI and Stanford Online launched an online course in partnership with the learning platform Apolitical to expand our education offerings for public sector employees. The training features instruction from multiple HAI experts and is designed to reach a broader audience of policymakers, including those in foreign governments. The course is currently provided free of charge and can be accessed on demand on the Apolitical platform.

CREDITS: AFORATIVE MEDIA / THE ASIA FOUNDATION



Society

In an expanded effort to engage in critical public interest conversations outside of policymakers, the policy team supports mission-driven civil society leaders in harnessing AI for social good, ensuring that the transformative potential of AI is aligned with the public interest.

Society Convenings

Inspiring Action: Identifying the Social Sector AI Opportunity Gap

This past year, HAI collaborated with Project Evident to conduct a nationwide survey for leaders in nonprofits and philanthropic organizations to assess their current use of and learning needs around AI. The findings were published in a working paper, "Inspiring Action: Identifying the Social Sector AI Opportunity Gap," which found that while nonprofits are using AI more than expected, there is still a notable gap between these organizations' interest in adapting AI technologies and the actual uptake.

This paper fed into a full day convening at Stanford on "Philanthropy and Equitable AI." The event brought together executive leaders from 44 philanthropic foundations with remarks by faculty members Alyce Adams, Michael Bernstein, Daniel E. Ho, and James Landay and a keynote address from Omidyar Network CEO Mike Kubzansky, who emphasized the need for greater philanthropic coordination on AI. Attendees engaged in community-building activities, table discussions, and field-building recommendations.

Designing Your Human-Centered AI Strategy: The Social Sector Cohort

The social sector plays a critical role in shaping the future of AI. In this program, leaders in social impact organizations, such as nonprofits and philanthropy, learn how to integrate AI into their organization's impact strategy, policies, and programmatic initiatives.

Research Centers & Labs

Stanford Digital Economy Lab (DEL)

The Stanford Digital Economy Lab is an interdisciplinary research group studying how digital technologies are transforming work, organizations, and the economy toward a world of shared prosperity. The lab leverages the expertise and interests of Stanford faculty, researchers, and students to initiate and support a wide range of projects on the most pressing topics related to the digital economy. As an integral part of Stanford HAI, DEL pursues a multidisciplinary approach that encourages dialogue and collaboration across the domains of economics, business, technology, policy, and beyond.

DEL has made significant strides over the past year. It welcomed its first Faculty Center Fellow and eight new postdoctoral fellows, all of whom are contributing to current and new projects in the lab's core areas of research:

Areas of Research

Measuring the Digital Economy: Creating better methods of measuring the health of an increasingly digital economy

AI & the Future of Work: Understanding the future of the workforce in a rapidly changing global economy

Digital Platforms and Society: Exploring how digital technologies can transform platforms and social media infrastructure to benefit society

MEDIA MENTIONS

What will artificial intelligence mean for your pay?
The Economist, November 16, 2023

ChatGPT provided better customer service than his staff. He fired them.
The Washington Post, October 3, 2023

Transforming the tax and finance functions for the future
Fast Company Executive Board, September 29, 2023

How can AI help you? What LLMs mean for the future of customer service jobs
Tech Brew, November 7, 2023

CREDITS: CHRISTINE BAKER

Thought Leadership

Economics of Transformative AI: In March 2024, DEL hosted a closed-door workshop, "What If We Succeed? The Economics of Transformative AI," bringing together AI leaders and economists to explore the implications of transformative AI—technology that could perform a significant share of tasks in the economy. The workshop addressed key questions about timelines, technological hurdles, and economic impacts, including productivity, wealth concentration, and global AI distribution. The event fostered crucial dialogue between technologists and economists, helping shape DEL's research agenda on the economic challenges and opportunities posed by transformative AI.

The Digitalist Papers: The Digitalist Papers explore how AI could reshape governance and legacy social institutions in profound and unpredictable ways. The inaugural volume was developed during AY 2023–24 and features a dozen original essays from leading thinkers across academia, industry, and civil society. To guide the writing process, DEL facilitated a full peer review of each essay and hosted an in-person workshop in May 2024, bringing authors together to refine their contributions. This project is funded by Frank McCourt and the Project Liberty Institute.

CREDIT: PATRICK BEAUDOUIN

Student Engagement

This past spring, DEL hosted the second iteration of its Stanford graduate-level course: *The AI Awakening: Implications for the Economy and Society*, which received several hundred applications from PhD, master's, and bachelor's degree students in the Graduate School of Business (GSB) and in Computer Science, Economics, and other departments. Each week, students learned from frontier researchers and industry leaders in technology, economics, and business, read the relevant research, and discussed implications. As a final capstone, students completed research proposals, policy proposals, and business plans. Topics included foundation models, large language models (LLMs), and generative AI; vision and robotics; work and employment; bias and explainability; AI and geopolitics; and a world without work. DEL was honored to be joined by esteemed guest speakers including Eric Schmidt, Mira Murati, Reid Hoffman, Anima Anandkumar, James Manyika, Shivon Zilis, Lawrence Lessig, and Daniel Susskind.

The 2023 course content remains live on Coursera. Thanks to a continued collaboration with the Stanford Center for Professional Development, DEL has made select videos from 2024 available for free to all learners on Stanford's YouTube channel. The published recordings of this course gained attention from wide audiences, accumulating over 100,000 views within the first two months.

CREDIT: PATRICK BEAUDOUIN

Cross-Campus Collaboration

DEL continues building relationships and research ties with other parts of the campus and faculty, fostering joint projects this fall and "bridge" researchers. In addition to its collaborations with GSB Professor Susan Athey on social media and misinformation, Computer Science Professor Michael Bernstein on AI agents/simulations, GSB Professor Chad Jones on AI and economic growth, and Sociology Professor Robb Willer on digital platforms and democracy, DEL added Computer Science Professor Diyi Yang and Stanford Center for Longevity Director Laura Carstensen to its network via joint postdoctoral fellows.

DEL Researchers

DEL's Advisory Group convened twice during the academic year. In October 2023, it discussed AI, social media and democracy, and milestones, timelines, and economic implications of transformative AI. In April 2024, members discussed an early-stage research agenda for transformative AI, as well as AI for scientific discovery.

Erik Brynjolfsson
Director

Christie Ko
Executive Director

Jerry Yang and Akiko Yamazaki Professor and Senior Fellow at HAI, Ralph Landau Senior Fellow at the Stanford Institute for Economic Policy Research (SIEPR), and professor, by courtesy, at the Stanford Graduate School of Business and Stanford Department of Economics

Faculty Fellow
Alex "Sandy" Pentland

Research Scientists
Sophia Kazinnik
Wang Jin
David Nguyen
Andrew Wang

Postdoctoral Fellows
Ruyu Chen
Ziv Epstein
Basil Halperin
Andreas Haupt
Christina Langer
Jiaxin Pei
José Ramón Enríquez
Philip Trammell
Gabriel Unger
Luca Vendraminelli

Students
Wajeeha Ahmad
Alma Andino Frydman
Alia Braley (UC Berkeley, & Stanford Project Scientist)

RESEARCH HIGHLIGHTS

GDP-B: Gross Domestic Product (GDP) is the sum of the value of all goods and services produced in a country in one year, where price is a proxy for value. GDP-B (B = Benefits) measures how much consumers benefit from goods and services, not just how much they pay. DEL's approach starts from basic principles of economics: Changes in well-being stem from changes in the economic surplus created by goods and services, rather than the money spent on them.

This past year, DEL expanded its large-scale experiments to measure consumer surplus across 700 items, including digital platforms, public goods, and the latest addition, measuring the value of "health." Additionally, the lab added new members (both on campus and abroad), including a new Stanford Research Scientist, Sophia Kazinnik, as well as faculty collaborators in other parts of the world, including at Tsinghua University and CEIBS, which are adapting its methodology and launching surveys in China.

DEL also assembled a group of experts from academia, industry, and governmental statistical agencies to join its GDP-B Advisory Committee to provide crucial feedback and diverse perspectives. Finally, the lab released a new website dedicated to this work, where forthcoming survey data will be published.

AI and the Future of Work: In a collaboration enabled by the Dieter Schwarz Foundation, postdoctoral researcher Christina Langer is working with Helmut Krcmar and Philipp Lergetporer of TUM Heilbronn to explore how automation and AI are changing the skills requirements for a variety of professions. In this multi-year engagement, Christina is splitting her time between Stanford and TUM Heilbronn, where she is conducting field research.

Misinformation Funding Experiment: Using survey experiments and decision-maker interviews, the lab examined how online misinformation is financed through advertising and found that most decision makers are unaware of their companies' ads appearing on misinformation sites and that consumers often react negatively when informed of these practices.

The monetization of misinformation is amplified by digital ad platforms, which algorithmically distribute ads across the internet. When informed about how platforms contribute to this process, decision makers express a preference for scalable, platform-based solutions to reduce misinformation financing. DEL proposed interventions to disrupt this financial support and counter the spread of misinformation. This research was published in Nature.

Generative AI for Enhanced Deliberation: DEL launched deliberation.io, a new platform designed to support large-scale discussions with customizable features aimed at improving public discourse and fostering consensus. It also completed a baseline study showing improvements in Americans' commitment to democratic practices, with upcoming pilots exploring how AI can aid in depolarization and collective action.

Center for Research on Foundation Models

The Center for Research on Foundation Models (CRFM) makes fundamental advances in the study, development, and deployment of foundation models. The center brings together a multidisciplinary group of faculty, students, researchers, and engineers spanning more than 10 departments at Stanford.

RESEARCH HIGHLIGHT

CRFM launched the Foundation Model Transparency Index in October 2023 as an annual initiative to assess leading foundation model developers. The 2023 study revealed an alarming lack of transparency among major companies, while the 2024 report showed considerable improvement with companies publishing new transparency reports via the Index. The index received widespread coverage in the media and widely consulted by policymakers across the United States, United Kingdom, Singapore, European Union, and United Nations.

MEDIA MENTIONS

Stanford researchers issue AI transparency report, urge tech companies to reveal more
Reuters, October 18, 2023

OpenAI Is Human After All; 'Sharing Is Caring,' Researchers Tell Model Developers
The Information, October 18, 2023

Why everyone seems to disagree on how to define Artificial General Intelligence
Fast Company, October 18, 2023

The world's biggest AI models aren't very transparent, Stanford study says
The Verge, October 18, 2023

We Don't Actually Know If AI Is Taking Over Everything
The Atlantic, October 19, 2023

Top AI Shops Fail Transparency Test: Stanford transparency index rates Meta, OpenAI, and others on 100 indicators
IEEE Spectrum, October 22, 2023

AI developers are failing on transparency, new index shows
Axios, October 24, 2023

Maybe We Will Finally Learn More About How A.I. Works
The New York Times, October 18, 2023

In AY 2023–24, the center also made progress across its three research areas:

Technical Research

Core technical contributions toward building, using, and understanding foundation models

- **Robust Distortion-free Watermarks for Language Models:** A best-in-class watermark for language models, helping developers to identify whether text is generated by their model
- **DoReMi:** Optimizing Data Mixtures Speeds Up Language Model Pretraining
- **Cybench:** A Framework for Evaluating Cybersecurity Capabilities and Risks of Language Models
- **AutoBencher:** Creating Salient, Novel, Difficult Datasets for Language Models
- **Holistic Evaluation of Text-To-Image Models:** A new framework for standardizing evaluation of text-to-image models, building on the substantial success of the HELM (Holistic Evaluation of Language Models) framework

Applied Research

Building foundation models and other artifacts that provide value in specific application domains such as medicine and law

- **OpenVLA: An Open-Source Vision-Language-Action Model:** A new approach for building vision-language-action models for robotics, achieving state-of-the-art performance on various robotics tasks
- **BioDiscoveryAgent:** An AI Agent for Designing Genetic Perturbation Experiments
- **Merlin:** A Vision Language Foundation Model for 3D Computed Tomography
- **MedAlign:** A Clinician-Generated Dataset for Instruction Following with Electronic Medical Records
- **LegalBench:** A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models

Societal Research

Shaping policy and clarifying the societal impact of foundation models to facilitate the responsible governance of this new technology

- **Considerations for Governing Open Foundation Models**
- **A Safe Harbor for AI Evaluation and Red Teaming**: A paper on the practices of leading foundation model developers, finding that none offer comprehensive protections for third-party researchers. The U.S. Copyright Office subsequently partially supported the researchers' position, stating that independent evaluation of AI models did not constitute violations of copyright law.
- **AIR-Bench 2024:** A Safety Benchmark Based on Risk Categories from Regulations and Policies
- **The AI Regulatory Alignment Problem**

CRFM Researchers

Percy Liang
Director, CRFM

Associate Professor of Computer Science, and, by courtesy, of Statistics; Senior Fellow, Stanford HAI

Rishi Bommasani
Society Lead, CRFM; PhD Candidate of Computer Science, Stanford University

David Hall
Research Engineering Lead, CRFM, Stanford University

Kevin Klyman
Master's Student, International Policy, Stanford University

Tony Lee
Research Engineer, CRFM, Stanford University

Yifan Mai
Research Engineer, CRFM, Stanford University

Julian Quevedo
Student, Computer Science and Mathematics, Stanford University

Josselin Somerville
Master's Student, Computer Science, Stanford University

Ivan Zhou
Master's Student, Computer Science, Stanford University

Additional HAI Centers, Labs, Initiatives, Partners, & Collaborators

Centers & Labs

- **Center for Research on Foundation Models (CRFM):** An interdisciplinary initiative at the Stanford Institute for Human-Centered Artificial Intelligence that makes fundamental advances in the study, development, and deployment of foundation models.
- **Center for the Study of Language and Information (CSLI):** CSLI serves Stanford faculty and students who are engaged in research involving computational, logical, and stochastic modeling of cognitive functions and processes.
- **Stanford Digital Economy Lab (DEL):** Pursuing a deeper understanding of the digital economy and its impact on the future of work and society

Initiatives

- **RAISE Health:** A joint initiative between Stanford Medicine and the Stanford Institute for Human-Centered Artificial Intelligence (HAI) to guide the responsible use of AI across biomedical research, education, and patient care
- **AI4ALL:** Empowering the next generation of AI change makers
- **Stanford Robotics Center:** Bringing together cross-disciplinary world-class researchers and industrial affiliates with a shared vision of robotics' future. Its unique collaborative facility supports large-scale innovative projects for transformative impact on people and the planet.

Research Partners

- **Center for AI Safety:** The mission of the Stanford Center for AI Safety is to develop rigorous techniques for building safe and trustworthy AI systems and establishing confidence in their behavior and robustness, thereby facilitating their successful adoption in society.
- **One Hundred Year Study of Artificial Intelligence (AI100):** A longitudinal study to examine and anticipate how the effects of artificial intelligence will ripple through every aspect of how people work, live, and play
- **Stanford Artificial Intelligence Laboratory (SAIL):** A center of excellence for artificial intelligence research, teaching, theory, and practice since its founding in 1963
- **Golub Capital Social Impact Lab:** Using digital technology and social science research to improve the effectiveness of leading social sector organizations
- **Open Virtual Assistant Lab:** Building LLM-augmented cognition systems (LLM-CSys) that combine the expressive language skills of LLMs with the interpretability and reliability of software systems, as well as connecting LLMs with databases, compilers, and executive functions to create assistants that are more trustworthy and helpful
- **Regulation, Evaluation, and Governance Lab (RegLab):** Stanford's RegLab builds the evidence base and technology for effective governance.
- **TUM Campus, Heilbronn:** Since fall 2018, the renowned Technical University of Munich is present at the Bildungscampus Heilbronn and offers a wide variety of services to its more than 500 students in five programs.

Collaborators

- **Black in AI Research (BlackAIR)**
- **Center for AI in Medicine and Imaging**
- **Center for Comparative Studies in Race and Ethnicity**
- **Center for International Security and Cooperation**
- **Cyber Policy Center**
- **Digital Civil Society Lab (DCSL)**
- **Hoover Institution**
- **Institute for Diversity in the Arts (IDA)**
- **John S. Knight Journalism Fellowships**
- **McCoy Family Center for Ethics in Society**
- **Office of the Vice President for the Arts**
- **Stanford Aging & Ethnogeriatrics Research Center (SAGE)**
- **Stanford Institute for Economic Policy Research (SIEPR)**

Industry Programs & Partnerships

Over the past year, HAI broadened its reach with industry partners while deepening existing collaborations, working within an expanding corporate ecosystem to guide decision makers and ensure that AI technologies are deployed with a strong focus on human-centered approaches.

Executive Education

Additional programming in AY 2023–24 included a Bangkok-based Executive Education program hosted with Accenture, an industry leaders conference on the periphery of the HAI at Five celebrations, and 11 executive breakfasts, featuring faculty including Julian Nyarko, Monica Lam, and Ge Wang. The HAI Startup Series continued to provide a unique setting for industry engagement with Stanford's AI-focused entrepreneurial community, showcasing 21 start-ups over seven sessions.

Industrial Affiliates Program

From September 2023 through August 2024, HAI significantly increased membership in the HAI Industrial Affiliate Program, adding new members including American Express, the Hanwha Group, LVMH, PwC, and SAP, from September 2023 through August 2024. These new relationships have expanded HAI's footprint in Europe and Asia and positioned the HAI Industrial Affiliate Program as the largest affiliate program at Stanford. Members include:

Research Collaborations & Faculty Engagement

HAI's industry research collaborations benefited more than 30 faculty members and engaged 50+ postdocs and graduate students.

Support has gone to a variety of fundamental and applied AI research domains spanning AI and society/economy, generative AI, responsible/trusted AI, foundation models, embodied AI, healthcare, fintech, and sustainability.

PROJECT HIGHLIGHTS

Building and Evaluating Multilingual MultiModal Models and Trustworthy Language Models with Conformal Guarantees

Offering new insights to fintech apps.

Diyi Yang

Assistant Professor, Computer Science, Stanford University; Faculty Affiliate, Stanford HAI

Holistic Evaluation of Language Models (HELM)

Adding financial-domain scenarios to HELM to cover the above financial-domain related aspects.

Percy Liang

Associate Professor, Computer Science, Stanford University; Director, CRFM; Senior Fellow, Stanford HAI

Network infrastructure for LLMs

Creating new solutions for key networking problems.

Balaji Prabhakar

VMWare Founders Professor, Computer Science and Electrical Engineering, Stanford University

RelBench: A Benchmark for Deep Learning on Relational Databases

A benchmark for deep learning on relational databases; a collection of realistic, large-scale, and diverse benchmark datasets for machine learning on relational databases.

Jure Leskovec

Professor, Computer Science, Stanford University

Deepening Interaction Between Retrieval and LMs

New mechanisms that facilitate deeper interaction between retrieval and LM components either in pre-training, fine-tuning, or post-training, in a way that reduces hallucination.

Dan Jurafsky

Professor, Linguistics and Computer Science, Stanford University

Stream of Search

Stream of Search (SoS) is a new framework that teaches language models to search for solutions by representing the search process itself in natural language. This framework:

- Targets naturalistic domains like mathematics and coding
- Creates a synthetic data generation pipeline that imbues language models with the ability to backtrack and self-correct, breaks down problems into sub-goals, and assigns values to intermediate states represented in natural language

Noah Goodman

Associate Professor of Psychology and Computer Science

Cross-Asset Time-Series Patterns for Investment: A Machine Learning Approach

A new approach that leverages the dependency of time-series patterns in multiple assets for optimal portfolio design in order to exploit lead-lag relationships and complex statistical arbitrage relationships between assets.

Markus Pelger

Associate Professor, Management Science & Engineering, Stanford University

About HAI

Stanford HAI brings together thought leaders from academia, industry, government, and civil society to shape the development and responsible deployment of AI.

People

Denning Co-Directors

John Etchemendy

Provost, Emeritus, and Patrick Suppes Family Professor in the School of Humanities and Sciences

John Etchemendy is a professor of philosophy and symbolic systems, teaching *Philosophy of Artificial Intelligence* to undergraduate students. During AY 2023–24, he co-led the HAI at Five: Celebrating 5 Years of Impact Conference in June 2024 and co-hosted U.S. Representative Anna Eshoo (CA-16) for the Unlocking Public Section AI Innovation: Next Steps for the National AI Research Resource seminar in October 2023. In his 35 years as a Stanford faculty member, he has served as director of the Center for the Study of Language and Information, chair of the Philosophy Department, senior associate dean for the School of Humanities and Sciences, and Stanford University's 12th provost.

James Landay

Anand Rajaraman and Venky Harinarayan Professor in the School of Engineering, Professor of Computer Science

James Landay was named Denning Co-Director for Stanford HAI during AY 2023–24. In February 2024, he received the SIGCHI Lifetime Research Award by the Association for Computing Machinery (ACM) for his career of leading cutting-edge research with his students. He co-led the HAI at Five: Celebrating 5 Years of Impact Conference in June 2024. During AY 2023–24, he focused his recent talks, including two HAI Research Seminars, on outlining a design process for human-centered AI: AI designers need to consider the needs and concerns of the AI's users, of the communities impacted by the AI's use, and of society at large.

Fei-Fei Li

Sequoia Professor, Computer Science Department

In September 2023, Fei-Fei Li was recognized as one of the most influential people in AI by TIME100 AI and received the Intel Innovation Lifetime Achievement Award. In October 2023, she spoke with Geoffrey Hinton, another recipient of the TIME100 AI award, during a masterclass. She co-hosted U.S. Representative Anna Eshoo (CA-16) for the Unlocking Public Section AI Innovation: Next Steps for the National AI Research Resource seminar in October 2023. She and President Emeritus John Hennessey conversed in November 2023 on her memoir *The Worlds I See: Curiosity, Exploration, and Discovery at the Dawn of AI*. She co-led the HAI at Five: Celebrating 5 Years of Impact Conference in June 2024 alongside Denning Co-Directors John Etchemendy and James Landay. She co-sponsored the AI + Health Conference in December 2023 alongside HAI Associate Director Curt Langlotz. During the 2024 Congressional Boot Camp in August, she spoke at the dinner keynote panel on the topic of AI's Global Impact of Democracy and Governance.

Fellows

HAI currently has over 220 fellows and affiliated Stanford faculty from across all seven schools. Each year, the institute aims to support the work of researchers who focus on the intersection of multiple disciplines. The HAI Fellows Program is made possible by generous donors like Martin and Millie Lau.

In AY 2023–24, HAI welcomed 30 new graduate and postdoctoral fellows from an array of disciplines, facilitating collaboration between engineers, social scientists, humanists, and others researching the future of human-centered AI. In 2025, HAI will welcome new Senior Fellows: Susan Athey, Michael Bernstein, Angele Christan, Dorsa Sadigh, and Melissa Valentine.

Senior Fellows

Russ Altman

Associate Director and Senior Fellow Kenneth Fong Professor and Professor of Bioengineering, of Genetics, of Medicine (General Medical Discipline), of Biomedical Data Science, and, by courtesy, of Computer Science

During AY 2023–24, Russ Altman helped organize the recruitment of several HAI Senior Fellows and served as a liaison between HAI and RAISE Health, a joint initiative between HAI and the School of Medicine. He co-hosted U.S. Representative Anna Eshoo (CA-16) for the Unlocking Public Section AI Innovation: Next Steps for the National AI Research Resource seminar in October 2023. He moderated a lightning session panel on the topic of "Implementation of AI in Society at Scale: How Do We Augment the Human Experience?" at the HAI at Five: Celebrating 5 Years of Impact Conference in June 2024. He is a faculty director of the AI100, a 100-Year Study of Artificial Intelligence. He is also the host of the podcast "The Future of Everything with Russ Altman."

Erik Brynjolfsson

Jerry Yang and Akiko Yamazaki Professor and Senior Fellow Director, Stanford Digital Economy Lab; Ralph Landau Senior Fellow, Stanford Institute for Economic Policy Research (SIEPR)

Erik Brynjolfsson spoke at the 2024 Congressional Boot Camp in August on the topic of AI, Automation, and the Future of Work. He also spoke at the HAI at Five: Celebrating 5 Years of Impact Conference in June 2024 on Implementation of AI in Society at Scale: How Do We Augment the Human Experience. During AY 2023–24, he hosted a number of workshops, including the New Measures of the Economy Workshop in March 2024, which explored new tools and techniques to measure modern economies, including the contribution of household production, people's health, and the environment; the Workshop on New Approaches to Characterize Industries: AI as a Framework and a Use Case to address the recent shifts in federal policy to invest in critical and emerging technologies; and the Workshop for Reinventing Digital Infrastructure for Civil Empowerment in May 2024 focused on understanding concrete pathways for AI-enhanced digital communications to promote human deliberation and enable productive collective action.

Michele Elam

Senior Fellow, William Robertson Coe Professor of Humanities, Department of English, Center for Comparative Studies in Race & Ethnicity, African & African American Studies

In AY 2023–24, Michele Elam taught undergraduate and graduate courses including *Black Mirror: A.I. Activism*, *Mixed-Race Politics and Culture*, *James Baldwin and American Culture*, and *Harlem Renaissance*. She was the lead PI, working closely with Stanford symbolic systems student Isabelle Levent, on the 2023 HAI Seed Grant recipient for her project, Stories of the Future: An AI and Film Convening, which brought together AI and HCI researchers, science fiction screenwriters, film directors, and producers to craft new depictions of AI, weaving in recent advancements and pressing philosophical/ethical concerns. She won the 2023 Norman Foerster Prize, awarded to the best essay of the year in American Literature for her piece: "Poetry Will Not Optimize; or, What Is Literature to AI?" in February 2024. Concurrently, she published "Digital Griots, Wampum Codes, and Choreo-Robotics: Artist-Technologists of Color, Reshaping the Public Sphere," which explored the impact of AI on philanthropy. Additionally, she spoke at several impactful conferences, including the Big Data and AI Paris Conference, and was the plenary keynote speaker at the Beyond Human-in-the-Loop: AI & "Human" Rights conference, the keynote speaker at the Generative AI: Leveraging AI for Economic Empowerment conference, a keynote panelist at the AI and Democracy summit on Renewing Democracy, Women Leading the Way for the opening of the Hillary Rodham Clinton Center, and the keynote speaker at the Leading Women Defined Summit on AI and Black Womanhood. She also joined the editorial board for the Cambridge Forum on AI, a global platform publishing and hosting conversations about emerging interdisciplinary work on AI.

Surya Ganguli

Associate Director and Senior Fellow, Associate Professor of Applied Physics and, by courtesy, of Neurobiology, of Electrical Engineering, and of Computer Science

Surya Ganguli co-hosted the 2023 HAI Signature Fall Conference, New Horizons in Generative AI: Science, Creativity, and Society, in October, which focused on AI in the sciences, creative disciplines, and society. His research spans the fields of neuroscience, machine learning, and physics, focusing on understanding and improving how both biological and artificial neural networks learn striking emergent computations. Over the academic year, he has been working on launching a new HAI lab with a research focus on the science of intelligence.

Daniel E. Ho

Senior Fellow, William Benjamin Scott and Luna M. Scott Professor of Law; Professor of Political Science; Professor of Computer Science (by courtesy); Senior Fellow, SIEPR; Faculty Fellow, Center for Advanced Study in the Behavioral Sciences (CASBS); Faculty Director, Stanford RegLab

Daniel Ho serves as the director of the Stanford RegLab, which partners with government agencies to design and evaluate programs, policies, and technologies that modernize government. He spoke at the September 2023 Hoffman-Yee Symposium to present his team's research project, which won a Hoffman-Yee research grant. In November 2023, he worked with the HAI Policy team to release "The AI Regulatory Alignment Problem" policy brief. He testified to the U.S. House Subcommittee on Cybersecurity, Information Technology, and Government Innovation on "Governing Artificial Intelligence" in December 2023 and worked on the HAI issue brief on "Considerations for Governing Open Foundation Models." In February 2024, he testified at the California State Senate Hearing on Securing California's Future: Artificial Intelligence and worked with HAI on the Safe, Secure, and Trustworthy AI EO Tracker that follows the implementation of the U.S. government's Executive Order 14110 on AI. In May 2024, he coauthored a new study, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," revealing the need for benchmarking and public evaluation of AI tools in law.

Curtis P. Langlotz

Associate Director and Senior Fellow, Professor of Radiology, Medicine, and Biomedical Data Science; Director, Center for Artificial Intelligence in Medicine & Imaging

As the director of the Center for Artificial Intelligence in Medical Imaging (AIMI Center) and associate director of Stanford HAI, Curtis P. Langlotz helps lead RAISE Health, a joint AI initiative of HAI and the School of Medicine. He also helped organize the 2023 AI + Health Conference, which convened experts and leaders from academia, industry, government, and clinical practice to explore critical and emerging issues related to AI's impact across the spectrum of healthcare. He serves as a reviewer for HAI-AIMI grant programs and the HAI Start-up series. He was elected President of the Radiology Society of North America in November 2023. He spoke at the Congressional Boot Camp about "Transforming Healthcare Through Innovation" in August 2024. In June 2024, he also gave a lightning talk at the HAI at Five: Celebrating 5 Years of Impact Conference on "Implementation of AI in Society at Scale: How Do We Augment the Human Experience?"

Percy Liang

Senior Fellow, Associate Professor of Computer Science, Stanford University; Director, Stanford Center for Research on Foundation Models

Percy Liang is the director of the Center of Research on Foundation Models (CRFM), a research center under HAI. He co-hosted the 2023 HAI Signature Fall Conference, New Horizons in Generative AI: Science, Creativity, and Society, in October, which focused on AI in the sciences, creative disciplines, and society. In October 2023, he coauthored the "Transparency Index on Foundation Models," which rated foundation model companies on their transparency. In September 2023, he coauthored the HAI Policy Briefs, "The Regulatory Alignment Problem," the "Foundation Models and Copyright Questions," and "Whose Opinions Do Language Models Reflect?" in November 2023. He was a speaker at the Congressional Boot Camp where he educated the attendees on "Understanding the Basics of Foundation Models" in August 2024.

Christopher Manning

Associate Director and Senior Fellow, Thomas M. Siebel Professor in Machine Learning, Professor of Linguistics and of Computer Science; Director, Stanford Artificial Intelligence Laboratory (SAIL)

Christopher Manning is the Director of the Stanford Artificial Intelligence Lab (SAIL) and a cofounder of HAI. In AY 2023–24, he coauthored the paper "Direct Preference Optimization: Your Language Model is Secretly a Reward Model," which was one of the two Outstanding Main Track Runner-Ups at NeurIPS 2023. He received the 2024 IEEE John von Neumann Medal for advances in computational representation and analysis of natural language. In August 2024, "GloVe: Global Vectors for Word Representation," which he coauthored, won the 10-year Test of Time Award at ACL. He spoke at the Joint Summit on Generative AI, a joint-conference between Stanford HAI and the University of California, Berkeley, focusing on how the state can best use generative AI to better serve the people of California and the impacts of GenAI on California and its workforce in May 2024. Also in May 2024, he coauthored a new study, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," revealing the need for benchmarking and public evaluation of AI tools in law.

Senior Fellows

Julian Nyarko

Associate Director and Senior Fellow, Professor of Law, Stanford University

Julian Nyarko became an Associate Director of Stanford HAI during AY 2023–24. He has participated and contributed to a number of HAI events, including the 2024 Congressional Boot Camp in August, where he educated attendees on "The Fuel of AI: Data (and its Perils)." He also released a new study, "What's in a Name? Auditing Large Language Models for Race and Gender Bias," in March 2024 that showed startling disparities in how the most popular LLMs give advice relating to individuals of a particular race or gender. In his study, the individual's identity was indicated by their first and last name. In May 2024, he published "Risk Scores, Label Bias, and Everything But the Kitchen Sink," a paper showing that in risk modeling, where AI researchers generally take a "more-is-better" approach to training data, using less information may often be preferable.

Rob Reich

Senior Fellow, McGregor-Girand Professor in Social Ethics of Science and Technology

Rob Reich's scholarship in political theory engages with the work of social scientists and engineers. His most recent books are System Error: Where Big Tech Went Wrong and How We Can Reboot (with Mehran Sahami and Jeremy M. Weinstein, 2021) and Digital Technology and Democratic Theory (edited with Lucy Bernholz and Hélène Landemore, 2021). He has testified before Congress and written widely for the public, including for the *New York Times*, *Washington Post*, *Wired*, *Time* magazine, *The Atlantic*, *The Guardian*, and the *Stanford Social Innovation Review*. He helped to create the global movement #GivingTuesday and serves as chair of its board. From April 2024 to January 2025, he is on public service leave to serve as Senior Advisor to the newly established United States Artificial Intelligence Safety Institute.

Ge Wang

Associate Director and Senior Fellow, Associate Professor, Center for Computer Research in Music and Acoustics (CCRMA); Department of Music and, by courtesy, of Computer Science, Stanford University

During AY 2023–24, Ge Wang became an Associate Director of Stanford HAI, where he focuses on helping everyone find more tools for critical reflection about AI and its use. Among his many initiatives, he piloted the HAI "Vodcast" series, *OffScript*, where he and HAI Research Director Vanessa Parli hosted thought leaders on the intersection of AI and ranging subjects. During AY 2023–24, he and Vanessa Parli hosted four *OffScript* episodes. Among other Stanford courses, he teaches *Music 356/CS 470: Music and AI (a critical making course)*. He gave a keynote performance at the October 2023 HAI Signature Fall Conference on New Horizons in Generative AI: Science, Creativity, and Society. He also served as a panelist during the AI + Education Summit, a conference co-hosted by HAI and the Stanford Accelerator for Learning in February 2024, on the topic of "Cutting Edge Work: Examples on AI and Education." During the closing keynote of the June 2024 HAI at Five: Celebrating 5 Years of Impact Conference, he and Celeste Betancur, a PhD candidate in Music, gave a presentation and live performance exploring AI and the creative process.

Amy Zegart

Associate Director and Senior Fellow, Morris Arnold and Nona Jean Cox Senior Fellow at the Hoover Institution; Professor, by courtesy, of Political Science; Senior Fellow at the Freeman Spogli Institute for International Studies

Amy Zegart is an internationally recognized expert in U.S. intelligence, emerging technologies, and global political risk management. She is the author of five books, including the bestseller *Spies, Lies, and Algorithms: The History and Future of American Intelligence*, which was nominated by Princeton University Press for the Pulitzer Prize. She leads Hoover's Tech Policy Accelerator and cochairs the *Stanford Emerging Technology Review* with Condoleezza Rice and Dean of Engineering Jennifer Widom. She serves on two U.S. Intelligence Community boards and frequently advises senior government officials on intelligence and national security matters. In AY 2023–24, she became an Associate Director of HAI and spoke at the Congressional Boot Camp keynote dinner on "AI's Global Impact on Democracy and Governance."

Faculty Fellows

Peter Norvig

Distinguished Education Fellow

Peter Norvig is the HAI Distinguished Education Fellow, where he helps to develop a human-focused AI curriculum and broaden access to education. A longtime innovator in the field, he co-wrote *Artificial Intelligence: A Modern Approach*, an introductory textbook used by 1,500 universities worldwide. He has taught hundreds of thousands of students through his courses on the online education platform Udacity.

Hariharan Subramonyam

Ram and Vijay Shriram HAI Faculty Fellow

Hariharan Subramonyam is an Assistant Professor at the Graduate School of Education and Computer Science (by courtesy). He is a Ram and Vijay Shriram Faculty Fellow at HAI. His research resides at the intersection of human-computer interaction (HCI) and the learning sciences. He studies methods to enhance human learning through AI by engaging in cognitively informed design practices, co-designing with learners and educators, and developing transformative AI-enabled learning experiences. Additionally, through this research, he contributes tools and methodologies that emphasize ethical considerations, responsible design practices, and human values in the creation of AI experiences.

Johannes Eichstaedt

Ram and Vijay Shriram, HAI Faculty Fellow

Johannes Eichstaedt is a computational social scientist jointly appointed as a Ram and Vijay Shriram HAI Faculty Fellow and Assistant Professor of Psychology. His lab in Computational Psychology & Well-Being is currently focused on using LLMs to develop interventions for well-being and mental health to augment therapy and support therapists. He also analyzes big social text data (Twitter, Facebook, Reddit) with natural language processing (NLP) and AI to understand and measure psychological processes for individuals and populations – for example, to build early warning systems for the opioid epidemic and mental health. Across these lines of work, his overall focus is on using AI and NLP to improve the well-being and health of the population.

Shriti Raj

HAI Faculty Fellow

Shriti Raj is a human-centered computing researcher. Her work focuses on augmenting health-related data reasoning and decision making. She is jointly appointed as a HAI Faculty Fellow and an Assistant Professor in the Department of Medicine Center for Biomedical Informatics Research. She received her PhD in information from the University of Michigan, where she was advised by Dr. Mark Newman. Her research develops and evaluates human-centered methods and tools to help people make health data and algorithms actionable.

Sandy Pentland

Center Fellow
Toshiba Professor of Media Arts & Science; Professor, Information Technology, MIT

Alex "Sandy" Pentland was appointed as a Center Fellow at HAI and as a Fellow and Faculty Lead for research on digital platforms and society at the HAI's Digital Economy Lab (DEL) during AY 2023–24. He also directs MIT's Human Dynamics Laboratory and the MIT Media Lab Entrepreneurship Program, co-leads the World Economic Forum Big Data and Personal Data initiatives, and is a founding member of the advisory boards for Nissan, Motorola Mobility, Telefonica, and a variety of start-up firms.

Policy Fellows

Jennifer King

Privacy and Data Policy Fellow, Stanford HAI

Rohini Kosoglu

Policy Fellow, Stanford HAI

Michael Kratsios

Distinguished Visiting Fellow, Stanford HAI; Managing Director of Scale AI and former CTO of the United States

Susan Rice

Distinguished Visiting Fellow, Stanford HAI; Former Domestic Policy Advisor to President Joe Biden; Former National Security Advisor for President Obama and U.S. Permanent Representative to the United Nations

Marietje Schaake

HAI International Policy Fellow; International Policy Director, Cyber Policy Center, Stanford University

Postdoctoral Fellows

Ruyu Chen
Stanford Digital Economy Lab Postdoctoral Fellow

Yuchen Cui
HAI Postdoctoral Fellow

Dapeng Feng
HAI Postdoctoral Fellow

Caroline Glidden
HAI Postdoctoral Fellow

Alex Goldmark
Journalist in Residence

Farnaz Jahanbakhsh
HAI Postdoctoral Fellow

Christopher Kelly
HAI Postdoctoral Fellow

Christina Langer
Stanford Digital Economy Lab Postdoctoral Fellow

Gabriel Unger
Stanford Digital Economy Lab Postdoctoral Fellow

Luca Vendraminelli
HAI Postdoctoral Fellow

Kaveh Waddell
John S. Knight Journalism Fellow

Luwen Wan
HAI Postdoctoral Fellow

Zhecheng Wang
HAI Postdoctoral Fellow

Staff

Russell Wald

Executive Director

Russell Wald oversees HAI's research, education, communications, administrative activities, industry programs, and policy and society hub. He collaborates with HAI's co-directors and faculty leaders to shape the strategic vision and human-centered mission of the organization. From 2020 to 2022, he served as HAI's first Director of Policy and later as Managing Director for Policy and Society. During this time, he established HAI's policy and society initiative, engaged with policymakers at the state, national, and international levels, created the impactful congressional boot camp, and coauthored several AI policy publications, including *Enhancing International Cooperation in AI Research: The Case for a Multilateral AI Research Institute* and *The Centrality of Data and Compute for AI Innovation: A Blueprint for a National Research Cloud.* Additionally, he serves as a member of the AI Index Steering Committee hosted by HAI.

Elena Cryst
Director of Policy and Society

Christie Ko
Executive Director, Stanford Digital Economy Lab

Anh Le
Director of Administration

Carolyn Lehman
Director of Media and Communications

Panos Madamopoulos-Moraris
Managing Director of Industry Programs and Partnerships

Vanessa Parli
Director of Research

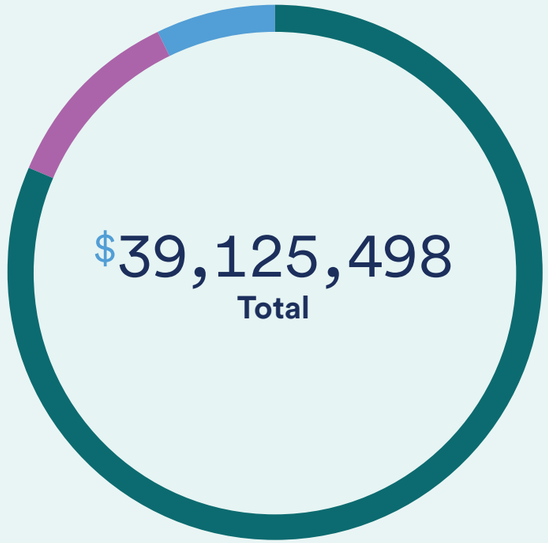
Stacey Sickels Boyce
Senior Associate Director of Development

HAI Advisory Council

The HAI Advisory Council consists of leaders in science, civil society, and business who provide HAI with expert advice on issues relevant to the institute's mission and programs. Reid Hoffman serves as chair of the council, along with vice chairs Steve Denning and Srinija Srinivasan.

Financial Data

Source of Income[1]



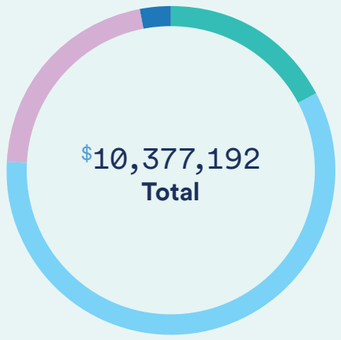
$2,720,930
Sponsored Research

$4,486,216
Endowment Payouts

$31,918,352
Gifts & Other Income

[1] Income does not include increases in endowment principal

HAI Research Grants Support



In FY24, HAI issued
$10.377M
in research grants

$1,795,800
Cloud Credit Grants

$6,075,000
Hoffman-Yee Grants

$2,206,392
Seed Grants

$300,000
AIM-HAI Partnership Grants

Expenditures

$30,694,201 **Total**

Stanford University
Human-Centered
Artificial Intelligence

Thank you

Thank you for following Stanford HAI's efforts to realize our vision of human-centered AI. We appreciate the support of our extended community and encourage you to contact us with any questions.

Support for HAI

Stanford HAI is grateful for the generosity of those who provide support to make its work possible. The impact of this philanthropic support extends beyond individual projects; it ripples wider as our researchers seek further discoveries, share their findings, and make strides toward new strategies for the enhancement of artificial intelligence and its role in shaping a better future for humanity.

The institute welcomes philanthropic contributions. Please contact Stacey Sickels Boyce, Senior Associate Director of Development, with interest or questions.

Email
sickels-boyce@stanford.edu

Office
650.497.3772

Mobile/WhatsApp
+1-410-991-4926