Computer Science and Artificial Intelligence Laboratory

The Computer Science and Artificial Intelligence Laboratory (CSAIL) pioneers approaches to computing that improve how people work, play, and learn. CSAIL serves the MIT community, the country, and society at large by creating a positive future enhanced by computer science through contributions of ideas, artifacts, and people, including:

- Integration of computing into the fabric of everyday life
- Acceleration of discovery in engineering, sciences, humanities, arts, and society
- Development of novel paradigms, models, techniques, software, and machines
- Creation of mathematical and scientific foundations for computing
- Training of students and postdoctoral researchers pursuing computer science research

Laboratory members conduct research in almost all aspects of computer science, including artificial intelligence (AI), the theory of computation, systems, machine learning, and computer graphics, as well as exploring revolutionary new computational methods for advancing health care, manufacturing, energy, and human productivity.

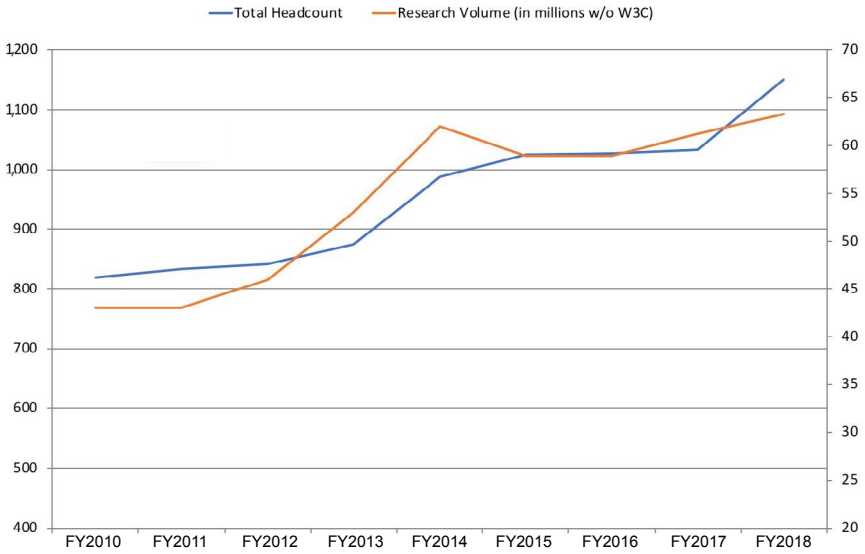
CSAIL has a long history of technological innovations that have affected how people interact and do business. CSAIL is known as the incubator for some of the greatest technological advances of the past 30 years, including the Internet, personal computing, mobile computing, open-source software, microprocessors, robotic surgery, and social networking.

CSAIL's current research addresses some of the grand challenges of the 21st century, including developing personalized learning, securing cyberspace, advancing health informatics, reverse engineering the brain, enhancing virtual reality, developing tools for scientific discovery, improving urban infrastructure, and ensuring the health of our environment. Computing is central to solving these challenges, and CSAIL contributes to making computing more capable by addressing fundamental algorithmic and systems questions at the core of computing and broadening the scope of computing to address important social challenges.

Key CSAIL initiatives currently under way include tackling the challenges of big data, developing new models for wireless and mobile systems, securing computers and the cloud against cyber attacks, rethinking the field of artificial intelligence, and developing the next generation of robots. Advanced software-based medical instrumentation and medical informatics systems to aid clinical decision making are being investigated. Advancements in biological research are also under way, including developments in the field of computational biology and application of machine learning to interpreting complete genomes and understanding gene regulation.

Over the past eight years, CSAIL has experienced 40% growth in headcount and 47% growth in research volume. With a total combined research volume of $66,392,336 for FY2018, CSAIL is the Institute's largest interdisciplinary research laboratory. The laboratory has seen steady growth in research volume during the past three years, with an approximate 3.7% increase between FY2017 and FY2018. CSAIL manages

over 350 active research awards and more than 110 principal investigators (PIs) with appointments across 11 MIT departments. In AY2018 545 graduate students had research assistant appointments in CSAIL, and we had 325 Undergraduate Research Opportunities Program (UROP) students.



CSAIL headcount and research volume, FY2010–FY2018.

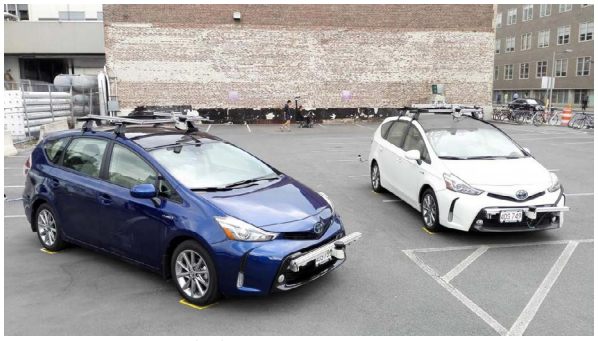
CSAIL research is sponsored by a large number of diverse sources, from US government contracts to the private sector. United States government sponsors include the Air Force Research Laboratory and the Air Force Office of Scientific Research, the Army Research Office, the Defense Advanced Research Projects Agency (DARPA), the Department of Defense Research and Engineering Enterprise, the Food and Drug Administration, the US Department of Education, the US Department of Energy, the Intelligence Advanced Research Projects Activity, the National Institutes of Health, the National Institute of Justice, the National Science Foundation (NSF), the US Navy (including the Office of Naval Research and the Naval Air Systems Command), and the Space and Naval Warfare Systems Center. US and international non-federal sponsors include Advanced Technology Laboratories, Boeing, BMW of North America LLC, the Ford Motor Company, the Foxconn Technology Group, the Intel Corporation, Jaguar Land Rover Limited, Lockheed Martin, the Microelectronics Advanced Research Corporation, the Nippon Telegraph and Telephone Corporation, the Northrop Grumman Corporation, Ping An Technology, the Qatar Computing Research Institute, Rakuten, Samsung Electronics, Suzhou Industrial Park, the Toyota Research Institute, and the Wistron Corporation. Other organizations sponsoring research include Aarhus University, the Battelle Memorial Institute, DSO National Laboratories, the Epoch Foundation, the Hong Kong University of Science and Technology, the Industrial Technology Research Institute, Nanyang Technical University, and the Singapore-MIT Alliance for Research and Technology.

Research Projects

CSAIL has many single- and multi-investigator projects as well as a number of virtual centers and large-scale projects. Examples of large-scale projects and collaborations are listed below.

Toyota-CSAIL Joint Research Center

Toyota established a collaborative research center with CSAIL and Stanford in 2015 to further the development of autonomous vehicle technologies, with the goal of reducing traffic casualties and potentially even developing a vehicle incapable of getting into an accident.



CSAIL car research demos.

Today, a car crash occurs every five seconds in the United States. Globally, road traffic injuries are the eighth-leading cause of death, with about 1.24 million lives lost every year. In addition to this terrible human cost, these crashes take an enormous economic toll. The National Highway Traffic Safety Administration has calculated the economic cost in the United States at about $277 billion per year. Putting a dent in these numbers is an enormous challenge—and this challenge is motivating the research of the Toyota-CSAIL Joint Research Center, established in September 2015. This center is a collaboration with the newly formed Toyota Research Institute, led by Gill Pratt.

Imagine that your car could tell you were having a bad day and turned on your favorite album to improve your mood. Imagine that your car could talk to your refrigerator, figure out that you're out of milk, and suggest where to stop on your way home. Imagine that your car knew you forgot to call your parents yesterday and issued a gentle reminder on the way home. And making that call was easy because you could turn the driving over to the car on a boring stretch of highway. These are just a few of the possibilities when cars and computer science are brought together. They are motivating research at the Toyota-CSAIL Joint Research Center.

The center's objectives are to advance AI and robotics research, develop a safe and intelligent car, and improve mobility and transportation by advancing the science of autonomy and machine intelligence. CSAIL researchers are working on (1) new tools for collecting and analyzing navigation data with the objective of learning from humans, (2) perception and decision-making systems for safe navigation, (3) systems that can handle difficult driving situations such as congestion, high-speed driving, and inclement weather, (4) predictive models that can anticipate the behavior of humans and vehicles, and (5) more intelligent user interfaces.

AY2018 projects conducted in the Toyota-CSAIL Joint Research Center were as follows:

- Geordi: A Driver's Assistant for Risk-bounded Maneuvering (PI: Brian Williams)
- Driver-Friendly Bilateral Control for Suppressing Traffic Instabilities (PI: Berthold Horn)

- Using Vision and Language to Read Minds (co-PIs: Nick Roy and Boris Katz)
- Uhura: A Driver's Personal Coach for Managing Risk (PI: Brian Williams)
- Predicting a Driver's State-of-Mind (co-PIs: Antonio Torralba and Wojciech Matusik)
- Exploring the World of High Definition Touch (co-PIs: Ted Adelson and John Leonard)
- Formal Verification Meets Big Data Intelligence to Address the Trillion Miles Challenge (PI: Armando Solar-Lezama)
- The Car Can Explain! (co-PIs: Gerry Sussman, Danny Weitzner, Hal Abelson, and Lalana Kagal)
- Crossing the Vision-Language Boundary for Contextual Human-Vehicle Interaction (co-PIs: Jim Glass and Antonio Torralba)
- Analysis by Synthesis Revisited: Visual Scene Understanding by Integrating Probabilistic Programs and Deep Learning (PI: Josh Tenenbaum)
- WiFi-Based Obstacle Detection for Robot Navigation (co-PIs: Dina Katabi and Daniela Rus)
- Drinking from the Visual Firehose: High-Frame-Rate, High-Resolution Computer Vision for Autonomous and Assisted Driving (co-PIs: Saman Amarasinghe, John Leonard, and Fredo Durand)
- Decision Making for Parallel Autonomy in Clutter (co-PIs: Daniela Rus and Sertac Karaman)
- A Parallel Autonomous Driving System (co-PIs: John Leonard, Sertac Karaman, and Daniela Rus)
- Uncovering the Pain Points in Driving (co-PIs: Ruth Rosenholtz, Fredo Durand, Bill Freeman, Aude Oliva, and Antonio Torralba)
- Simulation and Verification for Vision-in-the-Loop Control (PI: Fredo Durand)
- Tools and Data to Revolutionize Driving (co-PIs: John Leonard and Daniela Rus)
- Robotic Manipulation Data Engine (PI: Alberto Rodriquez)
- Dense, Freeform Tactile Feedback for Manipulation and Control (PI: Wojciech Matusik)
- Sensible Deep Learning for 3D Data (PI: Justin Solomon)
- All Terrain Mobility and Navigation (PI: Sangbae Kim)
- Inner Vision: Camera Based Proprioception for Soft Robots (co-PIs: Edward Adelson and Daniela Rus)
- A Safety Interlock for Self-Driving Cars (co-PIs: Daniel Jackson and Armando Solar-Lezama)

Wistron-CSAIL Research Collaboration

Good health—both mental and physical—is one of today's most pressing social and economic issues. A healthier population makes for a happier society and a more productive economy. Today, people are surrounded by an explosion of sophisticated and increasingly affordable information devices, from laptop computers, e-book readers, and smart glasses to mobile phones, smart watches, and health trackers. We monitor stock prices, weather forecasts, and traffic patterns through websites and apps; share our thoughts and experiences through emails, Facebook, and Twitter; and increasingly learn within online communities. These technologies open many new opportunities for improving how we live, work, and play. But how do they empower us and at what costs? Recent studies show that people consume 11 to 14 hours of technology each day. And this often involves multitasking, which in turn retrains our brains, reduces concentration, and increases stress. Finding ways to reduce stress and technology's negative impact on workforces is critical to our well-being in the future.

The multi-year Wistron-CSAIL research program focuses on rethinking how we compute and communicate in the digital age to ensure that health and well-being are at the core of our lives and that our use of technology accelerates this objective. Some of the questions we pose and the answers we seek include the following: How do we design the next generation of computers and communication systems to minimize our body's exposure to electromagnetic radiation? How should we rethink computer and communication architectures for sustainability? How can we develop systems that deliver appropriate lighting? How can we develop systems that re-engineer email? How do we develop algorithms that can help with information overload? How can we use computing and communication in support of individual and community well-being? How can we build computer and communication systems that are friendlier to our environment?

Our vision is to develop new computing and communication hardware and software platforms and supporting algorithms for modeling, controlling, and making decisions that will bring wellness to our use of technology. One thrust of the program focuses broadly on computer and communication platforms. The second thrust focuses on using these novel platforms to promote healthier living. Three projects are currently active in the Wistron-CSAIL Research Collaboration:

- Individual Prediction and Interpretation of Risk: Predicting Trajectories of Chronic Disease and Recovery (co-PIs: Polina Golland and Peter Szolovits)
- Smart Homes That Monitor Breathing, Heart Rate, and Life Quality (PI: Dina Katabi)
- Using Machine Learning to Build Better Clinical Support Tools (PI: John Guttag)

Suzhou Industrial Park–CSAIL Research Collaboration

AI is rapidly advancing and has the potential to be a great vector for positive change. On a global scale, new AI technologies will help us generate better insights into addressing some of our biggest challenges: curing diseases by better diagnosing, treating, and monitoring patients; understanding climate change by collecting and analyzing data from vast wireless sensor networks that monitor the oceans, the greenhouse climate, and plant conditions; improving governance through data-driven decision making;

eliminating hunger by monitoring, matching, and re-routing supply and demand; and predicting and responding to natural disasters using cyber-physical sensors. At an individual level, AI will offer opportunities to make our lives healthier, safer, more convenient, and more satisfying.

The multi-year research collaboration between Suzhou Industrial Park and CSAIL focuses on advancing the theoretical foundations of artificial intelligence. The questions we pose are centered around accelerating the processing of data, extracting semantic information from data, and developing integrated model-driven and data-driven optimization algorithms that will enable a variety of applications including smart cities and personalized health care. Four active projects are part of this collaboration:

- Inference for Machine Learning (PI: Polina Golland)
- Semantic Data Summarization with Coresets (PI: Daniela Rus)
- Software GPU for Machine Learning (PI: Nir Shavit)
- Geometric Optimization for Smart Cities (PI: Justin Solomon)

Ping An Research Projects

Many organizations today struggle to make intelligent use of the data they have collected and to determine how to structure their systems most efficiently. In addition, companies are continually looking for ways to not only automate but improve the customer experience.

To help achieve its goal of excelling in the financial services space, Ping An sought to partner with CSAIL to help address technical challenges. The Ping An collaborative began in 2014, structured as a three-year research engagement. Three projects were identified: Semantic Summarization for Financial Data, Speech Recognition, and Speaker Verification.

The semantic summarization project involves proposing a problem whereby there is a large set of database tables that contain data regarding customers, but tools do not exist for the required number of fields and records. To address this problem, we will build on our experience with summarizing data from mobile sensors such as smart phones, which are playing an increasingly important role in our lives and can be the source of useful data about the people carrying them. Our goal is to develop systems and algorithms that take large data streams and convert them into semantic summaries. The principal investigator for the project is Daniela Rus.

The second project focuses on Chinese speech recognition methods for customer support. Speech recognition capabilities would be useful to process what are currently millions of customer-agent call center interactions at Ping An. There is tremendous potential for speech technology benefits via spoken interfaces, including potential scenarios for future customer-computer spoken interactions via mobile devices. In the scope of this three-year project, however, the focus is on research from the Current Spoken Language Systems Group investigating the use of deep neural network–based methods for multilingual speech recognition. The project's principal investigator is Jim Glass.

Speaker verification, the focus of the third project, is the task of processing a speech recording and deciding whether or not it belongs to a putative speaker. For customer telephony applications, speaker verification is useful as a complementary verification method to existing approaches such as caller ID and PIN numbers. Speaker verification is closely related to speaker identification (or recognition), which seeks to determine the identity of a recording from a large candidate pool. There are several potential uses of speaker verification technology for call center customer support. Although the majority of these capabilities are directly relevant to customer-agent telephone-based call center communications, there are potential uses for customer-computer spoken interactions as well. Speaker verification technology complements existing methods for verifying customer identity. This project will use speaker diarization to automatically separate customer and agent speech turns during a single recorded dialogue. The principal investigator is James Glass.

The collaboration between Ping An and CSAIL has been very successful. The initial three-year term is expiring, and discussions are under way to extend or expand the program.

Qatar Computing Research Institute

In 2012 CSAIL began a seven-year, $35 million collaboration with the Qatar Computing Research Institute (QCRI) to address a wide range of research topics in computer science. Six years into the collaboration, we are currently pursuing eight projects, as follows.

- Arabic Speech and Language Processing for Cross-Language Information Search and Fact Verification (Senior Research Scientist Jim Glass, Professor Regina Barzilay, and Professor Tommi Jaakkola): This project aims to develop key technologies enabling cross-lingual search for verified facts and claims in multimedia content using questions posed in written and spoken language.
- Content-Adaptive Video Retargeting (Professor Wojciech Matusik): The goal of this project is to develop a complete system for delivering high-quality stereoscopic broadcast video, with a focus on real-time video of sporting events, particularly soccer.
- Database Management (Adjunct Professor Mike Stonebraker, Professor Samuel Madden, and Associate Professor Armando Solar-Lezama): This project focuses on building an end-to-end system to support the data discovery and preparation needs of data scientists, studying resource elasticity in online transaction processing database management systems, and using program synthesis for entity resolution and copy detection.
- Understanding Health Habits from Social Media Pictures (Professor Antonio Torralba): The major goal of this project is to understand food habits from social media images, including (1) using deep-image auto-tagging models to analyze food perception gaps, (2) learning a joint embedding space for recipes and food images, and (3) enabling vision-based estimations of population-level health from social media images.
- Understanding and Developing for Cultural Identities Across Platforms: Value-Driven Design Principles and Best Practices in a Qatari Context (Professor Fox Harrell): This project aims to (1) develop and deploy computational tools and

new techniques to understand use of virtual identity systems by Qatar and the Gulf Cooperation Council (GCC), (2) elicit and articulate best practices empowering Qataris to enact regional values and norms, and (3) develop GCC-specific novel applications.

- A Vertically-Integrated Approach to Resource-Efficient Shared Computing (Associate Professor Daniel Sanchez): This project aims to investigate an integrated node- and cluster-level architecture that provides both near-peak utilization and guaranteed performance in shared clusters, with a focus on QCRI workloads and infrastructure needs.
- Urban Data Analytics to Improve Mobility for Growing Cities in the Context of Mega Events (Associate Professor Marta Gonzalez): The goal of this project is to evaluate the impact of large-scale events and levels of accessibility based on travel times from different origins in Qatar as a means of sequentially proposing travel demand management strategies to mitigate traffic congestion during the 2022 FIFA (Fédération Internationale de Football Association) World Cup.
- Accurate Map Making Using Mobile Sensor Data (Professor Hari Balakrishnan, Professor Samuel Madden, Assistant Professor Mohammad Alizadeh, and Adjunct Professor David DeWitt): This project aims to develop accurate map-making techniques using crowd-sourced methods to overcome challenges related to creating and maintaining street maps, especially in a rapidly developing environment such as Doha, Qatar, leveraging data primarily from mobile phones and investigating current limitations resulting from sensor noise, outages, and data sparsity.

Centers and Initiatives

CSAIL Alliance Program

The CSAIL Alliance Program (CAP) is a gateway into the lab for industry and governmental institutions seeking a closer connection to the work, researchers, and students of CSAIL. The program provides organizations with a proactive and comprehensive approach to developing strong connections with all CSAIL has to offer. Leading organizations come to CSAIL to learn about our research, to recruit talented graduate students, and to explore collaborations with our researchers. Through this program, we are able to better provide our members with access to our latest thinking and our deep pool of exceptional human and informational resources. Overall, CAP supports the mission of CSAIL by connecting our researchers, students, and technological advances to organizations across the globe.

CAP uses a proactive and comprehensive approach to connect members to the entire laboratory—all 60 research groups spanning robotics, natural language processing, networks, databases, cryptography, web science, and more. CAP has two membership levels: affiliate and partner. Both include lab visits, access to the annual meeting, recruiting assistance, research briefings, and professional education discounts. Partner-level members, however, have more expanded options than affiliate-level members, with added access to research initiative meetings and custom faculty-led seminars as well as expanded recruiting options.

Currently there are more than 80 member companies, including global brands such as Apple, BASF, Google, Samsung, SAP, and Microsoft. Members are headquartered in North America, South America, Europe, and Asia and represent a wide variety of industry verticals.

CAP produces and manages online professional development courses in partnership with MIT Professional Education, MIT's Office of Digital Learning (ODL), edX, Get Smarter, and the Harvard Extension School. The table below provides a list of the courses offered to date along with brief course descriptions, numbers of offerings, and total enrollments. Overall course enrollment is now approaching 25,000 online learners, with total revenue to MIT of more than $850,000.

CSAIL Alliance Program Online Offerings to Date

Course title	Description	Number of offerings	Total enrollments
Artificial Intelligence: Implications for Business Strategy	Focuses on the organizational and managerial implications of AI technologies; offered in partnership with the Sloan School of Management	5	3,054
Tackling the Challenges of Big Data (not offered in FY2018)	Surveys state-of-the-art topics in big data, including data collection, data storage and processing, extraction of structured data from unstructured data, systems issues, analytics, visualization, and a range of applications	10	11,431
Tackling the Challenges of Big Data: Taiwan (not offered in FY2018)	Original course translated into traditional Chinese	1	1,296
Tackling the Challenges of Big Data: ILUMNO	Original course translated into Spanish and Portuguese and offered through ILUMNO in collaboration with universities in South America	1	375
Cybersecurity: Technology, Application & Policy	6-week online course providing a holistic look at cybersecurity technologies, techniques, and systems	8	3,199
HCI: Human Computer Interaction for User Experience Design	6-week course produced in partnership with Get Smarter; includes input from 8 CSAIL researchers who review a host of cutting-edge HCI concepts	2	126
Introduction to the Challenges & Opportunities of Big Data, the Internet of Things and Cybersecurity	Semester-long course combining our big data, IOT, and cyber courses; offered for credit through the Harvard Extension School	2	194
Startup Success: How to Start a Technology Company in 6 (Not So Easy) Steps (not offered in FY2018)	Discusses the lessons learned by Michael Stonebraker and Andy Palmer during their start-up endeavors over a 30-year period; lessons are distilled into 6 steps that any entrepreneur can follow to get a company going	2	359
Internet of Things: Roadmap to a Connected World	Introduces both the broad range of IoT technologies and the most recent developments in the space	7	3,897

SystemsThatLearn@CSAIL

The next decade will usher in a new frontier of sophisticated systems that perform complex "human-like" tasks, with complex inferences and predictions. With data being gathered from diverse sensors and mobile devices and computing power being spread across embedded devices and data centers, we will need new tools to realize the potential of learning systems. We are already seeing practical applications of these systems in areas such as autonomous vehicles and personalized health care that have the potential to transform industries and societies.

The goal of SystemsThatLearn@CSAIL is to accelerate the development of systems and applications that learn. We intend to accomplish this goal by combining our expertise in systems and machine learning to create new applications for understanding complex relationships unearthed by analyzing the avalanche of data available today.

At present, however, software systems that incorporate machine learning are difficult to build, deploy, and maintain. They require a large and highly skilled workforce. Unlike traditional enterprise systems, once built, they often require thousands of hours of ongoing, sometimes daily, maintenance to ensure that their predictions and behavior continue to be accurate and useful. Integrating machine learning systems into traditional enterprise architecture, testing, and deployment processes is too complex, partly due to organizational silos that exist between systems engineers and data scientists. Many problems in large-scale software systems involve optimizations that benefit from predictions, including scheduling, compilation, query planning, routing, data cleaning, and congestion control. Today, it is difficult to apply machine learning tools to design this type of system software.

Our approach to designing, training, and deploying "human-like" systems and applications will focus on four areas of investigation, as follows.

- Heterogeneous architectures: The data and features that drive learning in these systems and applications increasingly come from diverse distributed infrastructure, including phones, sensors, and other bandwidth- and power-impoverished endpoints. Thus, even acquiring data for learning may require adaptive allocation of computation over a heterogeneous infrastructure. Furthermore, the rise in heterogeneous hardware components, such as graphics processing units and many-core processors, suggests that a diversity of computational resources will be brought to bear.

- Predictable composition: Successfully designing and training machine learning systems for the desired task once data are available (i.e., programming at the level of learning components and reasoning about the behavior of such components) calls for skills and expertise that are not yet well supported or automated.

- Distributed execution: In terms of the underlying infrastructure, complex machine learning methods also demand considerable parallel resources if training is to be effective. Once trained, models may be deployed on massive parallel infrastructures (e.g., data centers) or may have to be reduced and distributed back to heterogeneous components where they can be used as needed (e.g., mobile devices), requiring new distributed algorithms and execution frameworks.

- Seamless integration of training and deployment: Many machine learning solutions today are trained and deployed in well-separated phases of training and testing (deployment), but this will change. Learning will increasingly become an ongoing, integrated process. The tighter integration of learning and computer systems offers exciting possibilities in terms of new capabilities but requires us to overcome challenging hurdles pertaining to programming abstractions, maintenance and monitoring, analysis, and performance guarantees.

In addition to building better systems for machine learning, we believe that our focus on deployability of models will help us advance machine learning itself by developing new models designed to further the above democratization goals while still providing excellent prediction accuracy. We expect many new tools and practices to be developed.

The large, multi-PI SystemsThatLearn@CSAIL initiative aims to accelerate the development of this next generation of systems. The primary focus is on developing a common infrastructure, specifically in the form of software that includes new theoretical advances:

- Tools that help data scientists and engineers understand their models, scalably train them, monitor results, and retrain the models efficiently
- A focus on efficiently deploying models in distributed settings, reusing and redeploying models, and creating development environments effective in training and deployment
- Development of heterogeneously deployable models or lower fidelity models that can run on sensors or smartphones as well as on more powerful servers
- Tools for statistical monitoring and performance prediction in cases where machine learning is used to understand the performance of complex systems
- Tools and methods to implement and run systems that learn over an untrusted infrastructure

SystemsThatLearn@CSAIL is led by Professors Sam Madden and Tommi Jaakkola and includes 37 CSAIL researchers. It is structured as an industry consortium. We launched the initiative on March 29, 2017, with six founding members: BT, NOKIA Bell Labs, Microsoft, Salesforce, ScotiaBank, and Schlumberger. BASF, EY, JPMorgan Chase, and ElementAI joined in FY2018.

FinTech@CSAIL

Throughout 2017–2018, the laboratory ramped up and prepared for the launch of the new FinTech@CSAIL initiative. Financial technology is disrupting many aspects of financial services, banking, insurance, and other industries. Not only will infrastructure and operations be disrupted, but new technologies, business models, services, and even industries will be launched. Financial technology holds promise for verified transaction systems such as block-chain systems as well as for technologies involving AI, security, data analytics/value extraction, machine learning, trust verification, risk management, and privacy advances.

The financial sector has many unique attributes, and at the core of a company's success in this sector are trust, security, value, and efficiency. Current technology roadmaps are not perceived as providing sufficient guidance for financial technology, and new players and technologies are constantly emerging. The shifting demands of customers are evident and pose both risk and opportunity. To stay competitive and stay ahead, companies need access to innovation, thought leadership, new technologies, and high-caliber talent.

FinTech@CSAIL will bring together industry, thought leaders, innovators, academics, developers of disruptive technologies, and start-up companies that are reinventing global financial services. We will work closely with industry partners in leveraging innovation from cutting-edge research to develop the next generation of impactful technologies that will open up new business models, broaden access, allow new data insights, and improve security.

The breadth of research at CSAIL uniquely positions the laboratory to address a wide variety of challenges in the space. FinTech@CSAIL will include 15 CSAIL researchers who have pioneered the fields of secure computation, machine learning, artificial intelligence, data analytics, and risk management. The goal is to advance the state of the art in collaboration with select industry partners as a means of addressing the most difficult problems facing the finance industry today.

Through our faculty's rigorous research coupled with our tradition of collaborating with industry, FinTech@CSAIL will address relevant business problems with a long-term vision. Additionally, FinTech@CSAIL will draw from across the laboratory as well as other focused research initiatives in Cybersecurity@CSAIL and SystemsThatLearn@CSAIL, producing a powerful collaboration to address the challenging issues that are emerging. We anticipate many opportunities for direct, active collaboration and knowledge sharing though events, projects, and directed research.

By leveraging the research ecosystem at CSAIL, we will work to address areas such as the following: new approaches to efficient shared public ledger systems and digital currencies; technologies to provide secure multi-party computation as well as secure and private data extraction; advanced data analytics and their applications to risk management and prediction; scalable, trusted systems; and machine learning and AI system security. We will also address speech recognition applications, legacy system security, efficient processing and automation of tasks, anonymization of data and privacy, secure sharing of data sets, and new lending and payment technologies.

FinTech@CSAIL will launch on July 18, 2018. The initiative is led by Professors Andrew Lo, Silvio Micali, and Shafi Goldwasser. Founding members include NASDAQ, Ant Financial, Citigroup, Ryan Software, the London Stock Exchange Group, and Ripple Labs. State Street Bank and the Canadian Imperial Bank of Commerce have also joined.

Cybersecurity@CSAIL

Cyber systems cover communications, banking, data processing, purchasing, power/energy infrastructure, transportation, defense—nearly every aspect of our lives. Consequently, in addition to becoming more frequent, cyber-based attacks are becoming more devastating. The present weaknesses in both hardware and software continue to

threaten not only the confidentiality of private data and the integrity of data at large but also the availability of the critical operating systems organizations use to support internal operations, manage assets, and secure logistics. Today these cybersecurity challenges rise across virtually all industry sectors, and organizations are dealing with ever-increasing numbers of attacks.

Through Cybersecurity@CSAIL, we are not just designing technology for specific tasks but working toward solutions for the entire security spectrum. We approach security from all sides: programming languages, software verification, computer architecture, crypto, systems, and policy. Our goal is to create security "by default" and remove program error as a source of vulnerability. We are establishing new theoretical and practical foundations of secure computing that integrate security into the design process.

Our objective is to design protocols that make attacks more difficult, retain function despite attacks, and allow a system to recover quickly from an attack. Cybersecurity@CSAIL intends to maintain an interdisciplinary focus that brings together industry and government thought leaders, along with MIT faculty, researchers, and students, to conduct research across the security spectrum in the areas of hardware, software, encryption, and theory. Specifically, we will address challenges related to operating system security, secure codes, hardware designs for optimal security, defense tools, cloud security, multi-party protocols, and usability of encrypted data.

The Cybersecurity@CSAIL industry consortium launched in March 2015. Our industry partners, which include Boeing, Raytheon, BBVA, Akamai, and State Farm, provide valuable perspectives on the challenges faced across several industry verticals. New projects undertaken through this initiative in FY2018 include the following:

- Cybersecurity in Multi-Robot Networks (Professor Daniela Rus)
- Security Monitors for Industrial Control Systems (Howard Shrobe)
- Adversarial Fraud Prevention (Una-May O'Reilly)

Internet Policy Research Initiative

Communication and information networks are a fundamental infrastructure for our increasingly digital economy and society. Technologists and policymakers both play key roles in this area, yet they approach issues from different perspectives. This can lead to policy-making that is not fully informed and to misdirected research efforts. Because of society's reliance on this critical infrastructure, there is a pressing need to bridge the gap between the technical and policy communities.

The mission of the Internet Policy Research Initiative (IPRI) is to work with policymakers and technologists to increase the trustworthiness and effectiveness of interconnected digital systems. This is accomplished through targeted engineering and public policy research, various educational programs geared toward students and policymakers, and outreach programs to build policy communities that facilitate communication, education, and information exchange. Core research efforts cover six distinct categories, as follows.

- Cybersecurity: IPRI's cybersecurity research focuses on security issues related to communication networks and software systems as they affect the economy and society as a whole. The work covers encryption policy, core infrastructure, and a joint interdisciplinary project (CSAIL and the Sloan School of Management) measuring corporate cyber risk in a secure and private way that allows data to be gathered, computed, and shared without the individual data source being divulged.

- Privacy: IPRI maintains a strong focus on privacy policy, including its critical role in trustworthiness. Research projects include understanding the privacy impacts of human/computer interaction, developing and strengthening privacy infrastructure, evaluating the relationships between security and privacy, and assessing the complex terrain of citizens' rights and state authority. Privacy is a key international focus area for the initiative, with various projects coming out of the Privacy Bridges report and IPRI's Keys Under Doormats report.

- Networks: IPRI research examines the communication networks that support the economy and society and includes topics related to assessing the reliability of services, reducing the cost of content delivery, measuring network performance, facilitating the disclosure of mechanisms for an open Internet and spectrum licensing, and analyzing the future of Internet architecture.

- Critical infrastructure: The digital systems that control critical infrastructure in the United States and most other countries are easily penetrated and architecturally weak. These vulnerabilities have been evident for a long time, but policymakers and system operators have tended to focus on short-term fixes and tactical improvements. The IPRI research stream focuses on developing and supporting short- and long-term recommendations that are applicable to critical infrastructure in the United States and abroad.

- Artificial intelligence policy: As machine learning tools become increasingly valuable in systems outside advertising and research, they stand to become increasingly prevalent in traditionally regulated spaces such as health care, finance, transportation, and employment. While each of these contexts involves certain legal and policy guarantees that decisions made can be understood and that the information on which they are based is robust, the technical novelty and complexity of machine learning tools complicate existing means for doing so. IPRI is helping bridge the gap between them by identifying the technical state of the art in interpretability and robustness in machine learning systems and applying those insights to evaluation and reconception of the legal, regulatory, and design requirements of regulated, accountable systems. Bringing together technical and policy scholars, IPRI provides a venue to help align technical and policy research; determine the most promising approaches to policy-applicable, accountable systems; and inform a broader research agenda encompassing multiple contexts in which machine learning systems will proliferate in the economy and society.

- Internet experience: IPRI research on the Internet experience includes the study of Internet governance, reflection on the role and evolution of information and communications technologies in society, and analysis of how key sectors use these technologies. Additional research tracks focus on socially linked data,

accountable systems, and the security of autonomous vehicles. As part of this innovative technical research, IPRI's App Inventor group provides a platform for developing policy-aware applications.

World Wide Web Consortium

The World Wide Web Consortium (W3C) was founded at MIT in 1994 by the inventor of the web, Tim Berners-Lee. W3C is responsible for developing and maintaining the standards that make the web work and for ensuring the long-term growth of the web. More than 400 member organizations, including most of the world's leading technology companies, are working to enhance the capabilities of web documents and create an open web platform for application development. This platform would be available across a wide range of devices, enabling collaboration and sharing of data and information.

In recent years, a great many factors (people, devices, bandwidth, policy decisions) have extended the reach of the web in society. Video, social networking tools, user-generated content, location-based services, and web access from mobile devices are transforming many industries, including mobile communications, television, publishing, the automotive and entertainment industries, games and gaming, and advertising. This transformation has led to greater demands on W3C and other organizations to build robust technology that meets society's needs in areas such as privacy, security, accessibility, and multilingual content.

Core Technology Focus

W3C standards define an open web platform for application development that has the unprecedented potential to enable developers to build rich interactive experiences powered by vast data stores that are available on any device. Although the boundaries of the platform continue to evolve, industry leaders speak nearly in unison about HTML5 as the cornerstone for this platform. But the full strength of the platform relies on many more technologies that W3C and its partners are creating, including cascading style sheets, scalable vector graphics, web open font formats, real-time communications, the Semantic Web Stack, and a variety of application programming interfaces. The platform continues to grow, and the W3C community, in turn, is growing to meet the demand.

With the completion of HTML5, there are many new areas of focus. Publicly noted security breaches have resulted in unprecedented attention to improving web security. The growth of e-commerce has focused new attention on standardizing payment and e-commerce approaches. In addition, with the Internet of Things (IoT) arriving, our Web of Things project aims to address semantic interoperability to prevent the IoT from driving silos at the application level. With immersive technologies, there is a strong emphasis on web solutions for virtual reality and augmented reality.

The demand is also driving W3C to expand its agenda and the size of its community. W3C launched Community and Business Groups in 2011. After seven years, roughly 10,000 people are participating. By making it easier for people to participate, W3C has increased the relevance and quality of its work and brought more innovators to the table for pre-standards and standards track work.

Industry Impact and Broadening the Set of Participants

In recent years, web technology is not only used by consumers and companies for information sharing, but increasingly the web is the delivery mechanism for companies to deliver their services. Examples of that include telecommunications (where web access is a key service), entertainment (which is increasingly delivered over the web), publishing (whose standards organization, the International Digital Publishing Forum, recently merged into W3C), and retail and financial services (both impacted by an increase of payments on the web). This has resulted in a diversification of the W3C membership and also has enriched the technical agenda to address new issues that arise. For example, web browser companies, credit card companies, and other financial technology stakeholders are using W3C to streamline web payments through browsers.

Other Activities

During the past year, CSAIL leadership spearheaded several other efforts that benefited the broader MIT community. For example, CSAIL launched a new website and a vision and rebranding project and took part in the AI and the Future of Work Congress, the MIT Digital Healthcare Conference, and various other forums. Also, the laboratory engaged in discretionary and research fundraising and explored further financial opportunities. Space planning and refurbishment efforts were initiated, and a new CSAIL lobby design was implemented.

Research Highlights

In addition to large-scale collaborative projects and other center research, numerous individual and multi-investigator projects are under way. A sampling of this work appears below.

Stefanie Mueller: Dynamic Materials for 3-D Printing

In light of the current sustainability challenge and the increased speed at which consumers buy and trash products, we are investigating how to develop products that can update themselves. Rather than fabricating new versions, we are developing methods to change the appearance and function of existing objects without requiring new materials.

One of our solutions includes a method for changing the color of a 3D-printed object even after fabrication. Our method is based on photochromic inks that can switch their appearance from transparent to colored when exposed to light of a certain wavelength. The color remains even when the object is removed from the light source. The process is fully reversible, allowing users to recolor the object as many times as they want. To switch between different colors, we use a multi-color pattern with one color per voxel across the surface of the object. When recoloring the object, our system locally activates only those voxels with the desired color and turns all other voxels off.

Our method has implications across many different areas, ranging from clothing that can change its appearance several times during the day depending on the occasion (e.g., business meeting versus casual lunch) to personalization in the sharing economy (e.g., recoloring a car's interior depending on the current user). It also has important implications for safety (changing the highlighted color depending on the background) and camouflage.

Arvind: Big Data Analysis within Solid State Disks

The economic value of big data is directly related to our ability to analyze it. Many complex analysis problems, such as detection of terrorist networks based on interactions on social media or detection of cancerous mutations in a patient's genome, can be formulated as graph problems. Solving large graph problems requires tremendous resources and is usually done in the cloud using a cluster of expensive server-class machines that can accommodate all of the data. We have been exploring whether some of these graph problems can be solved efficiently without bringing all of the data into dynamic random-access memory (DRAM) simultaneously.

In a recent paper we showed that, for several important graph problems, a big NAND flash store coupled with a small amount of DRAM and hardware accelerators can outperform a 16- to 32-node cluster machine. When packaged commercially, such a system would resemble a solid state disk that can be plugged into a PC. In the future, doctors may be able to perform complex analyses of a patient's genome in their office using a single machine that costs less than $1,000, as opposed to the currently required cluster of servers costing over $10,000.

Boris Katz: Language Proficiency

We developed a novel approach for determining learners' second-language proficiency based on patterns of eye movements during reading. Our approach provides eye-tracking-based English proficiency scores that reflect the extent to which a learner's gaze patterns in reading are similar to those of native English speakers. We showed that our scores correlate strongly with standardized English proficiency tests. We also demonstrated that gaze information can be used to accurately predict the outcomes of such tests. By deriving proficiency as an automatic byproduct of eye movements during ordinary reading, our approach offers a valuable new tool for second-language proficiency assessment.

It is currently estimated that more than 1.5 billion people are learning English as a second language worldwide. Their learning progress is commonly evaluated via standardized language proficiency tests (e.g., the Test of English as a Foreign Language) that are taken by millions of language learners every year. Despite their ubiquity, traditional approaches to language proficiency testing have several drawbacks; for example, the tests are prepared manually and require extensive resources, and their validity can be undermined by test-specific training, prior knowledge of the evaluation mechanisms, and plain cheating resulting from unauthorized access to test materials.

In our approach, we determine language proficiency from analyses of eye movements during reading of free-form English text. Our framework does not require test takers to prepare for the test or to perform any handcrafted linguistic tasks; rather, they are simply required to attentively read an arbitrary set of sentences. To the best of our knowledge, this work is the first to propose and implement such an approach, yielding a novel language proficiency evaluation scheme that relies solely on ordinary reading. We have demonstrated the effectiveness of the approach via score comparisons with standardized English proficiency tests. Our results confirm the promise of our methodology to reliably measure language proficiency.

Building on this work, we plan to study how specific aspects of linguistic proficiency (for example, specific grammatical constructions or knowledge of specific words) manifest themselves in specific patterns of eye movement during reading. Insights from such analyses will enable us to deepen our understanding of the interactions among language proficiency, reading patterns, and linguistic properties of text. This will be the first step toward using eye-tracking technology to provide learners with informative feedback on their test performance and help them improve their language skills.

Mohammed Alizadeh, Hari Balakrishnan, and Sam Madden: Making Accurate Maps from Aerial Imagery

Mapping road networks is both expensive and labor intensive. In many areas of the world, including Southeast Asia, Africa, and Latin America, digital maps are missing roadways, while in other areas they have errors in complex intersections and topologies. Accurate and up-to-date maps are especially important given the popularity of location-based mobile services and the impending arrival of autonomous vehicles. Several companies are investing hundreds of millions of dollars on mapping the world; however, despite this investment, map providers receive tens of thousands of error reports each day. For example, in Doha (Qatar), Google Maps has been missing entire subdivisions for years. ”

High-resolution aerial imagery provides a promising avenue to automatically infer a road network. We have developed a new method, RoadTracer, that automatically constructs accurate road network maps from aerial images. RoadTracer uses an iterative search process guided by a decision function implemented with a convolutional neural network to derive the road network graph directly as an output. RoadTracer is more accurate than prior methods using convolutional neural networks to classify each pixel as a road or not a road, and it works well even in the case of cities for which there are no training data. We evaluated RoadTracer on 15 cities (after training on 25 other cities) and found that, at a 5% error rate, it correctly captured 45% more junctions than prior methods.

CSAIL PhD students Favyen Bastani and Songtao He are participating in this project, which is being conducted in collaboration with Sofiane Abbar and Sanjay Chawla at QCRI in Qatar. Ongoing work related to the project includes using Global Positioning System (GPS) data to improve accuracy when imagery is occluded or missing and to determine directionality and intersection topologies, as well as using machine-assisted editing tools to augment human efforts in map making.

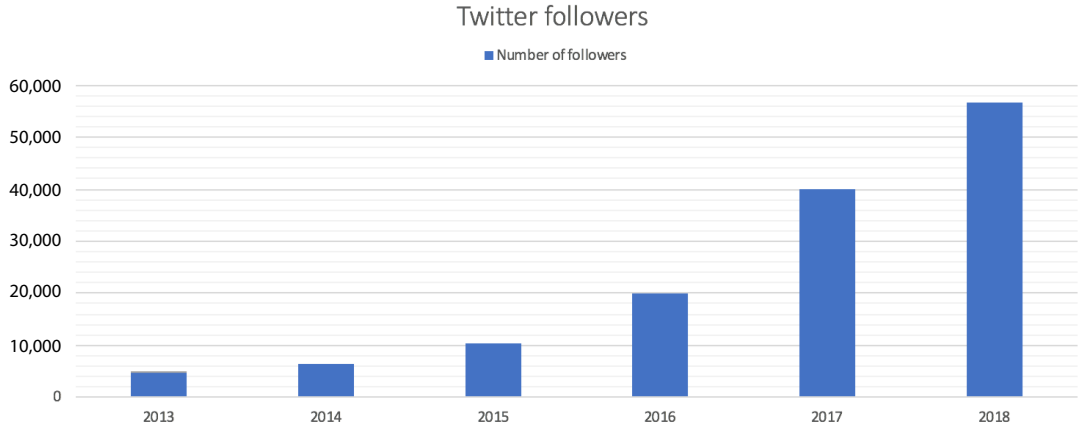
Laboratory-Sponsored Activities

CSAIL Outreach

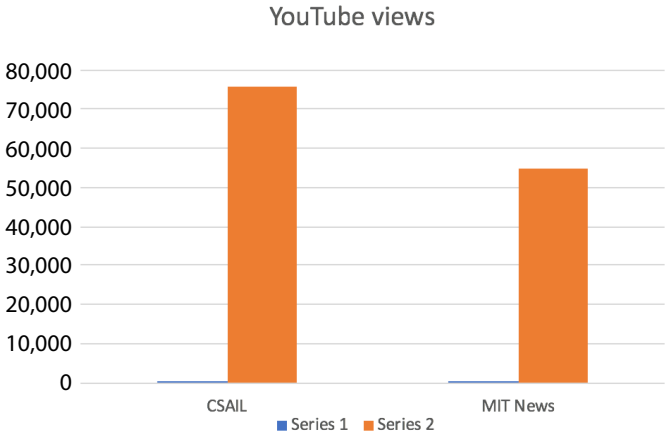
CSAIL regularly encourages the online community to submit questions about computer science and academia to its researchers in a series of Reddit “Ask Me Anything” (AMA) sessions. CSAIL’s AMAs have spurred approximately 7,000 comments and questions as well as more than 250,000 page views.

Media Outreach

CSAIL has a combined following of 114,000 users across Twitter, Instagram, Facebook, and YouTube channels. The laboratory's YouTube videos have an average viewership of 75,821. There has been a 940% increase in CSAIL media coverage and viewership over the last five years. In 2018, viewers have watched 1.26 million minutes of CSAIL video.



Growth in Twitter followers.



YouTube views comparison.

Distinguished Lecture Series

The Dertouzos Lecture Series has been a tradition since 1976, featuring some of the most influential thinkers in computer science. Two speakers presented Dertouzos lectures in AY2018: Yann LeCun (New York University and Facebook AI Research), "How Could Machines Learn as Efficiently as Animals or Humans?" (November 1, 2017), and Tim Berners-Lee (CSAIL and World Wide Web Consortium), "From Utopia to Dystopia in 29 Short Years" (May 2, 2018).

In 2017, CSAIL inaugurated Hot Topics in Computing, a community speaker series convening experts in computing to discuss perceptions and problems associated with the proliferation of computation and machines. Speakers included:

Josh Tenenbaum, "AI and Common-sense Reasoning" (November 29, 2017)

Adam Belay, Srini Devadas, and Joel Emer (with Sam Madden as moderator), "Spectre and Meltdown: Recent Security Flaws" (January 17, 2018)

Silvio Micali, Nickolai Zeldovich, and Neha Narula, "Blockchain Technologies and Applications" (February 21, 2018)

Daniel Weitzner and Ilaria Liccardi, "Off the Privacy Cliff: Facebook, Putin and the Prospects for Individual Rights Online" (April 22, 2018)

Organization

Professor Daniela Rus continues in her role as director of CSAIL. The director's duties include developing and implementing strategies designed to keep CSAIL growing and evolving, fundraising, determining laboratory policies, and examining promotion cases.

CSAIL's leadership team includes two associate directors, a chief operating officer (COO), and the executive cabinet. These leaders are appointed by the laboratory's director and assist her with her duties. Professors Daniel Jackson and Charles Leiserson are the current associate directors. Professor Leiserson also serves as chief operating officer, providing leadership and strategy in terms of how we conduct our operations and events, which enables the director to allocate more time to strategic planning. Victor Zue holds the role of director of greater China relations, managing oversight of various important CSAIL international contracts and contract negotiations.

The CSAIL executive cabinet meets twice per month to review and advise the director on policies, processes, and activities within the laboratory. AY2018 executive cabinet members included Hal Abelson, Ted Adelson, Saman Amarasinghe, Randall Davis, Daniel Jackson, Charles Leiserson, Sam Madden, Wojciech Matusik, Ronitt Rubinfeld, Daniela Rus, Nir Shavit, Bruce Tidor, and Victor Zue.

The CSAIL enterprise services team manages lab operations. Seven units (administrative assistants, the CSAIL Alliance Program, communications, finance, human relations, special projects, Infrastructure Group) report to the CSAIL COO on all operational matters. Carmen Popovic is the assistant director for administration. Jack Costanza is the assistant director for infrastructure, overseeing information technology infrastructure and user support, building operations, and communications. Lori Glover is managing director of the CSAIL Alliance Program. Victoria Palay is the senior manager of special projects.

Bruce Tidor is the current space czar, overseeing the space committee and managing allocation of space within CSAIL. The space committee implements facility improvements that will enhance the quality of the environment for the laboratory's faculty, staff, and students. The space committee also includes Jack Costanza.

Faculty Appointments, Leaves, and Promotions

Three new assistant professors joined CSAIL in AY2018: Adam Belay, Julian Shun, and Tim Kraska.

Faculty members on leave during 2017–2018 included the following:

Hal Abelson (sabbatical, fall)

Regina Barzilay (sabbatical, spring)

Tim Berners-Lee (sabbatical, spring)

Tommi Jaakkola (sabbatical, spring)

Daniel Jackson (sabbatical, fall)

Frans Kaashoek (sabbatical, spring)

Aleksander Madry (sabbatical, fall)

Wojciech Matusik (sabbatical)

Robert Miller (sabbatical, spring)

Gerry Sussman (sabbatical, spring)

David Gifford (leave without pay)

William Freeman (leave without pay)

Shafi Goldwasser (leave without pay)

Stefanie Jegelka (parental leave, spring)

John Leonard (on partial leave with the Toyota Research Institute)

Russ Tedrake (on partial leave with the Toyota Research Institute, spring)

Fox Harrell was promoted to professor in AY2018, and Adam Chlipala, David Sontag, and Vinod Vikuntanathan were awarded tenure.

Awards and Honors

Our faculty and staff received many awards and honors this year. Some examples are:

Mohammed Alizadeh: Association for Computing Machinery (ACM) Special Interest Group on Data Communication (SIGCOMM) Rising Star Award, ACM SIGCOMM Best Paper Award, Sloan Foundation Research Fellowship

Hari Balakrishnan: ACM Special Interest Group on Management of Data (SIGMOD) Test of Time Award

Regina Barzilay: Association for the Advancement of Artificial Intelligence Fellowship, ACL Fellowship, MacArthur Fellowship, Ruth and Joel Spira Award for Excellence in Teaching

Tamara Broderick: Alfred P. Sloan Research Fellowship, Amazon Research Award, Army Research Office Young Investigator Program Award, Google Faculty Research Award, Jerome H. Saltzer Award, NSF CAREER Award

Michael Carbin: NSF CAREER Award

Constantinos Daskalakis: Google Faculty Award

Srini Devadas: Charles A. Desoer Technical Achievement Award

Alan Edelman: Institute of Electrical and Electronics Engineers (IEEE) Fellow

Shafi Goldwasser: ACM Fellow, BBVA Foundation Frontiers of Knowledge Award, Carnegie Mellon University Honorary Doctorate of Science and Technology

Daniel Jackson: ACM Special Interest Group on Software Engineering (SIGSOFT) Outstanding Research Award, Arthur C. Smith Award, Martin Luther King Leadership Award

Stefanie Jegelka: DARPA Young Faculty Award, Joseph A. Martore Award for Exceptional Contributions to Education in IDSS, Sloan Research Fellowship

David Karger: ACM SIGCOMM Networking Systems Award

Dina Katabi: ACM Prize in Computing, ACM SIGCOMM Test of Time Award

Tim Kraska: Very Large Databases Conference Early Career Research Contribution Award, VMware Systems Research Award

Butler Lampson: Royal Society Foreign Membership

Tom Leighton: ACM SIGCOMM Networking Systems Award, Marconi Prize

Charles Leiserson: ACM SIGCOMM Networking Systems Award

Barbara Liskov: IEEE Computer Society Computer Pioneer Award

Tomas Lozano-Perez: ACM Fellow

Sam Madden: ACM SIGMOD Test of Time Award

Alexander Madry: Presburger Award

Silvio Micali: ACM Fellow, BBVA Foundation Frontiers of Knowledge Award

Robert Miller: Richard J. Caloggero Award

Stefanie Mueller: ACM Doctoral Dissertation Honorable Mention Award, ACM Special Interest Group on Computer-Human Interaction (SIGCHI) Outstanding Dissertation Award, MIT EECS (Electrical Engineering and Computer Science) Outstanding Educator

Ronald Rivest: BBVA Foundation Frontiers of Knowledge Award, National Inventors Hall of Fame inductee

Daniela Rus: American Academy of Arts and Sciences Fellow, IEEE Pioneer in Robotics and Automation Award, American Society of Mechanical Engineers Nyquist Lecturer

Julie Shah: IEEE Robotics and Automation Society Early Academic Career Award in Robotics and Automation

Peter Shor: IEEE Eric E. Sumner Award, International Centre for Theoretical Physics Dirac Medal, IEEE Information Theory Society Best Paper Award

Michael Sipser: ACM Fellow

Justin Solomon: Professor Amar G. Bose Research Fellowship

Michael Stonebraker: ACM SIGMOD Test of Time Award

Russ Tedrake: *International Journal of Robotics Research* Paper of the Year Award

Vinod Vaikuntanathan: DARPA Young Faculty Award, Harold E. Edgerton Faculty Achievement Award

Nickolai Zeldovich: ACM Special Interest Group in Operating Systems (SIGOPS) Mark Weiser Award

Key Statistics for Academic Year 2018

Faculty: 99 (16% women)

Paid UROP participants: 325 (41% women)

Postdoctoral researchers: 85 (12% women)

Principal research scientists: 10 (60% women)

Research staff: 47 (23% women)

Graduate students: 443 (22% women)

Visitors: 78 (21% women)

Master of engineering students: 102 (31% women)

Administrative, technical, and support staff: 90 (61% women)

Lecturers: 2 (0% women)

Senior research scientists: 4 (25% women)

Daniela Rus
Director