Computer Science and Artificial Intelligence Laboratory

The Computer Science and Artificial Intelligence Laboratory (CSAIL) studies computing and its application to finding solutions to many of the most challenging problems of our lives, our work, and our world. We employ a long-term lens to engineer innovative solutions in an effort to unlock the secrets of human intelligence, extend the functional capabilities of machines, explore human/machine interactions, and enhance the human experience. Our mission is to be an innovative and passionate community driving the future of computing and making the world a better place.

With approximately 50 research groups working on hundreds of diverse projects, CSAIL researchers focus on finding innovative ways to make systems and machines operate faster, better, safer, easier, and more efficiently for the benefit of humanity. Our projects fall into three areas of inquiry:

- Artificial intelligence—understanding and developing both living and artificial systems capable of intelligent reasoning, perception, and behavior
- Systems—discovering new principles, models, metrics, and tools of both hardware and software-based computer systems
- Theory—understanding the mathematics of computation and its wide-ranging, real-world consequences

CSAIL has a long history of technological innovations that affect how people interact and do business. CSAIL is known as the incubator of some of the greatest technological advances of the past 30 years—true life-changers, 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 heath 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, as well as broadening the scope of computing to address important social challenges that confront us. Current research explores cloud computing, mobile computing, the next generation of computing devices, autonomy, intelligence, and the application of sensor technology to traffic congestion, medical monitoring, and climate and environmental observations. Robotic locomotion and the human-robot interface are being investigated, as well as advanced software-based medical instrumentation and medical informatics systems to aid clinical decision making. Advancements in biological research are also under way, including developments in the field of computational biology and the application of machine learning to the interpretation of complete genomes and the understanding of gene regulation.

CSAIL research is sponsored by a large number of diverse sources, from US government contracts to the private sector. US government sponsors include the Air Force Research Laboratory and the Air Force Office of Scientific Research, the Army Research Office, the Defense Advanced Research Project Agency (DARPA), Department of Defense Research and Engineering, the US Department of Education, the Department of Energy, the Intelligence Advanced Research Projects Activity, the National Institutes of Health, the National Institute of Justice, the National Science Foundation, the Navy (including the Office of Naval Research and Naval Air Systems Command), and the Space and Naval Warfare Systems Center. US and international industrial sponsors include Boeing, Ford, Foxconn, General Electric, Intel Corporation, Lockheed Martin Advanced Technology Laboratories, Microelectronics Advanced Research Corporation, Nippon Telegraph and Telephone Corporation, Nokia, Northrop Grumman Corporation, Qatar Computing Research Institute, Quanta Computer, Inc., Shell, Siemens, and State Farm Insurance. Other organizations sponsoring research include Aarhus University, Battelle Memorial Institute, Delta Electronics Foundation, DSO National Laboratories, Epoch Foundation, Industrial Technology Research Institute, Nanyang Technical University, and the Singapore–MIT Alliance.

Research Projects

CSAIL comprises many single- and multi-investigator projects as well as a number of virtual centers and large-scale projects. The large-scale projects and collaborations include:

Quanta/Qmulus Project

The T-Party project is a 10-year, $45.5 million research collaboration with Quanta Computer, Inc. The project has just completed its eighth year and is currently funding four major research areas.

> *Cloud computing technologies*: a new multi-core operating system (sv6) that explores a possible contract between application implementers and kernel developers (professor Frans Kaashoek); a system (Poirot) that can efficiently check for past intrusions into a web application to facilitate intrusion recovery (professor Nickolai Zeldovich); a cloud-based caching infrastructure (Pequod) for web sites that cache the results of time-consuming database queries (professor Robert Morris); a scalable cloud-based database as-a-service (DBaaS) focusing on estimating how several concurrent databases on a cloud service will interact (professor Samuel Madden); a six-node prototype cluster of field-programmable gate array (FPGA)-based flash controllers, demonstrating the effectiveness of hardware-software co-design (professor Mithal Arvind); and novel data structures that allow efficient client manipulation of large server-side data streams (professor Stephen Ward)
>
> *Multimedia*: putting “human intelligence on demand” behind interactive applications by exploring crowd-powered user interfaces (professor Robert Miller); continued evolution of a probabilistic framework for natural language understanding and dialogs, as well as an organic framework for system development to support interface development for wide-ranging applications, including entertainment, tourist information, and healthcare (Drs. James Glass

and Stephanie Seneff); development of low-cost hardware implementation for speech recognition (professor Anantha Chandrakarsen); stereoscopic post-production of images (professor Wojchieh Matusik); and a cloud-enabled interactive robot (professor Seth Teller)

Healthcare: exploring cloud-based environments and applying computer vision, graphics, and machine learning techniques to develop medical analytics; demonstrating their usefulness in a variety of applications, including medical telemetry, non-invasive detection of pause rate, blood flow, and motion (professors John Guttag, Fredo Durand, and William Freeman)

Education technology: authoring tools for online audiovisual presentations (Professor Durand); crowd computing for education (Professor Miller); “see one, do one, teach one”—an improved pedagogy for online teaching of engineering subjects (Dr. Christopher Terman); and (in engineering education) developing curricular application programming interfaces (APIs) (Professor Ward) and human language technology for massive open online courses (professor Victor Zue)

Qatar Computing Research Institute

In late June 2012, MIT (represented by CSAIL) and the Qatar Foundation (represented by the Qatar Computing Research Institute, or QCRI) signed a seven-year, $35 million agreement to collaborate on a wide range of research topics in computer science. The objective of the agreement was to create connections between researchers in Qatar and those at MIT, and advance the state of the art in various areas of computer science.

During the current year, 11 projects involving 14 Principal Investigators (PIs) were initiated, covering the following areas:

- Arabic language technology: generating breaking news using Twitter and learning Arabic dialectal morphology (professor Regina Barzilay), multilingual speech informatics (James Glass), and opinion mining and its incorporation into an Arabic multimodal dialogue system (Dr. Stephanie Seneff)
- Data analytics: MAQSA—interactive social analytics for news (Sam Madden); and Data Tamer—a scalable data integration system (professor Michael Stonebraker)
- Distributed systems: cloud platform for end-to-end video retargeting (professor Wojchieh Matusik)
- Cyber security: Rapid prototyping of hardware security solutions (professor Mithal Arvind)
- Scientific computing: computational methods for high resolution genetic mapping (professors David Gifford and Tommi Jaakkola); and computational neuro-modeling of visual memory (Dr. Aude Oliva and professor Polina Golland)
- Social computing: mobile linked application (mobile app) development for disaster management (Dr. Lalana Kagal and professor Hal Abelson)

During spring 2012, the joint collaborating oversight committee met to refine the funding model in two ways. First, to achieve greater and longer-term impact, the project will fund a set of thematic umbrellas involving multiple PIs. Second, in addition to the current collaboration model involving PIs from both sides, a new development model has been introduced in which CSAIL PIs will form teams to help QCRI develop areas currently not being pursued but that are deemed important for a world-class computing research organization. Since April, 14 pre-proposals have been submitted, resulting in five projects (four under the collaboration model and one under the development model) being selected in late June to go forward over the next three years.

Basic Research in Interactive Computing

The Basic Research in Interactive Computing project is a six-year, $5.7 million research collaboration sponsored by Hon Hai/Foxconn Technology Group. This research collaboration is completing its last year of phase two and is in the process of being extended to a third phase. The project is currently funding four major research avenues in several areas of computer science, ranging from networking to human-computer interactions, computer graphics, vision, and theory. The research is predicated on the belief that computers and information devices are fast becoming interactive; they interact with other computers, with their environments, and, above all, with humans. Each form of interaction adds a new dimension to the challenge of modeling and understanding the behavior of computer systems as well as the task of building and using these systems.

During the current year, the collaboration supported four multi-year projects and six PIs in the following research areas:

- Advanced hand-tracking and gesture-based interaction (professor Randall Davis)
- Factor-analysis–based speech and language analysis (James Glass)
- Activity modeling and recognition: using phones as sensors (professor Daniela Rus)
- 3D Capture and Display for Mobile Devices (professors Fredo Durand, William T. Freeman, and Antonio Torralba)

A total of eight Foxconn engineers in two shifts visited CSAIL during the 2012–2013 academic year to facilitate technology transfer and to receive training for advanced research and development in computer science. In addition, video presentations for the four projects were made via video conferencing from CSAIL to Foxconn in Taiwan in December 2012 to summarize the research results.

The MIT Angstrom Project: Universal Technologies for Exascale Computing

The CSAIL-led Angstrom team was one of four teams selected by the Defense Advanced Research Project Agency for funding under the Ubiquitous High-Performance Computing program in 2010. The team is charged with the task of rethinking computing and creating a fundamentally new computing architecture to meet the challenges of computing in the 2020 timeframe. The CSAIL team is the only university-led team of the four, consisting predominantly of university researchers, the majority of whom are MIT faculty. Angstrom is a strongly interdisciplinary program, involving faculty from MIT's CSAIL, Microsystems Technology Laboratories, and the Research Laboratory of Electronics, and faculty from the University of Maryland Department of Electrical and Computer Engineering. Project Angstrom's goal is to create the fundamental technologies necessary for future extreme-scale computers. Extreme-scale computers face several major challenges, the most difficult four being the energy efficiency challenge, the scalability challenge, the programmability challenge, and the dependability challenge. To address these challenges, Angstrom is reexamining every layer of system design and interfaces including circuits, hardware architecture, operating systems, runtime software systems, compilers, programming languages, and applications. Angstrom adds self-awareness to each of these layers in order to improve energy efficiency and dependability while retaining programmability and scaling to thousands of processors.

Self-aware systems will not only make computers smarter, they could also prove essential for managing ever more complex computers in the future. Over the past decade, computer engineers have added more and more basic computing units, called cores, to computers. Today's computers have eight to 16 cores, but future machines will use anywhere from dozens to thousands of cores. That would make the task of splitting up computational tasks among the cores, which programmers now do explicitly, nearly impossible. In AY2012, an overall design of a 1,000-core Angstrom processor was completed. In the past year, several prototype processors were built in that exploit many of the design innovations in Angstrom. These are described below.

Energy-Aware Processor

We have prototyped a single "tile" of Angstrom that has energy-aware features. The tile consists of a processor, an adaptive cache, and energy monitoring circuits. The cache size can be changed via software instructions and the cache is therefore adaptive. The cache memory is implemented using Static Random Access Memory cells that can be run at very low voltage, allowing the entire tile to run at very low voltage. The processor and the cache energy can be monitored through software; i.e., the amount of energy consumed by any set of instructions by the processor and by the cache is measured by energy monitoring circuits and these energy quantities can be read by software. This energy-aware processor will be able to run applications with a strict energy budget by augmenting the applications to constantly monitor energy usage.

Execution Migration Machine

In the Angstrom Project, we have prototyped hardware-level instruction-granularity thread migration in a 110-core chip multiprocessor. Implemented in 45nm application-specific integrated circuit technology, the chip occupies 100 square millimeters and is currently in the fabrication stage. With a custom stack-based instruction set architecture to enable partial context migration, when there is no network congestion, this implementation provides end-to-end thread migration latency of four cycles between neighboring cores with a minimum thread context, and 33 cycles between the farthest cores with a maximum context. To supplement a remote-cache-access–based shared memory paradigm, the cores learn a thread's data access patterns and migrate threads automatically. Through register-transfer-language–level simulation, we have demonstrated that thread migration can reduce on-chip data movement by up to 14 times over a remote-cache-access architecture.

In-Network Coherence Processor

Multicore parallelism is imperative, as the drive to keep apace with Moore's Law is requiring application performance improvement despite transistor scaling limitations. With the emergence of the many-core era, scalable coherence and on-chip communication networks are crucial for shared memory processors. While snoopy coherence is prominent today, directory-based coherence is the de facto choice for scalability to many cores as snoopy coherence relies on ordered interconnects that do not scale. In recent years, researchers have proposed in-network snoopy coherence on a scalable on-chip network. A 11 mm by 13 mm 36core chip prototype, fabricated in IBM 45nm silicon on insulator technology, details a practical in-network coherence architecture and design. The ordered mesh network interconnect architecture contains distributed in-network global ordering and an adapted coherence protocol closely designed with the network, connected to 36 Freescale Power Architecture cores and Cadence on-chip memory controllers. In addition, the NoC is compliant with standard bus protocols allowing for system-on-chip busses to be interchanged easily.

The Ford-MIT Alliance

The Ford-MIT Alliance, an Institute-wide initiative, was established in 1997. In 2013, the Alliance began a new fourth phase that will renew automatically each year for the foreseeable future. The Alliance is the Institute's longest running large-scale commitment from industry, and represents a significant acknowledgment by the Ford Motor Company of MIT's relevance and impact on its research. Since 1997, the Alliance has funded more than 150 projects of varying duration and budgets across the Institute, with a total investment by Ford to date of more than $47 million.

The Ford-MIT Alliance research portfolio has a strong connection to CSAIL's research areas. CSAIL faculty are currently engaged in research projects under the Ford-MIT Alliance in four areas:

- Algorithms for driving history analysis
- Novel speech interfaces
- Crowd computing
- Open source applications for automobiles

CSAIL hosts a testbed of two Ford vehicles, a 2011 Ford Explorer and a 2012 Ford Focus, for use in these projects.

Intel Science and Technology Center in Big Data

MIT PIs Sam Madden and Michael Stonebraker continue to head the Intel Science and Technology Center in Big Data based at CSAIL. Professors Madden and Stonebraker are leading a team of 20 researchers (from Brown University, MIT, Portland State University, University of Tennessee, University of California Santa Barbara, and University of

Washington) in a project to build new software tools to process and manage massive amounts of data. Specifically, the center is focused on new data management systems and new computer architectures that together can help users process data that exceeds scale, rate, or sophistication of data processing such as existing systems provide. The center will develop new technologies in a number of areas, including data-intensive scalable computing, machine learning, and computer architecture and domain sciences (e.g. in genomics, medicine, oceanography, imaging, and remote sensing).

This was the center's first year of full operation, achieving a number of milestones:

- Hired several postdoctoral associates and one full-time research staff member, and established a website (http://istc-bigdata.org) with an active blog and social media presence
- Collected a set of test data sets and starting building benchmarks in the social media, graphic, and genomics areas
- Developed a number of new systems and algorithms for big data, including a major new interactive visualization system, a highly scalable similarity search engine, a new high-throughput transactional database system, and a number of extensions designed to optimize the performance of systems that themselves optimize the placement and execution of linear algebra and database-style queries over matrices
- In close collaboration with Intel engineers, began efforts to understand how next-generation hardware innovations, such as many-core chips, non-volatile random-access memories, and reconfigurable hardware impact the design of data processing systems
- Initiated SW-Store, a system for interactive processing of massive online data streams

Robotics Callenge

The DARPA Robotics Challenge was inspired by the Fukushima nuclear disaster and other situations in which human-like mobility and dexterity is required, but which are too hazardous for humans to go into physically. (Other examples of such situations include the Deepwater Horizon oil rig blowout and the Yarnell, AZ, wildfires.) Through the competition, DARPA aims to accelerate the development of remotely operated humanoid robots that can walk and climb over complex terrain within disaster zones (rubble, stairs, ladders), perform useful tasks that require handling objects (levers, pumps, valves) using tools (wrenches, drills, saws), and even climb into and drive ordinary vehicles. The goal of the competition: for the next Fukushima-type disaster where human bodies cannot help—even with protective gear due to intense heat and radiation—it should be possible to send robots that can take steps to prevent or mitigate disaster. Assuming that humans in the disaster zone were evacuated quickly leaving behind tools and vehicles, these implements should be available for use by the robots.

CSAIL quickly formed a team in spring 2012 and wrote a proposal to participate as a "Track B" team, which involved writing software that runs on a government-provided humanoid robot. DARPA reviewed the proposal favorably in July 2012 and funded CSAIL's team starting in October 2012.

First, CSAIL initiated a substantive collaboration between two research groups: professor Seth Teller's group (focusing on perception and human–robot interactions) and professor Russ Tedrake's group (focusing on real-time planning and control). We repurposed much of the infrastructure that had been developed as part of MIT's participation in the 2006–2007 DARPA Urban Challenge. We also developed new methods of controlling a humanoid robot at a level between teleoperation and full autonomy. Our strategy requires the human operator to make a high-level plan to complete the mission at hand, then break that plan into a series of smaller pieces, each of which can be executed autonomously by the robot. However, before the operator grants the robot permission to proceed autonomously, the robot must convince the operator, through a visual display of its interpretation of its surroundings and its planned actions, that it will be able to execute the operator's instructions.

Throughout the summer and fall of 2012 and the spring of 2013, we fleshed out our novel interface, perception, networking, planning, and control methods and implemented them in a DARPA-mandated simulation environment. This work was put to the test in June 2012 as part of the Virtual Robotics Challenge that DARPA used to select which of the 126 total software teams would be provided with a humanoid robot; MIT's team was one of the seven teams selected.

We are expecting delivery of our humanoid robot in July 2013, and will spend the next five months preparing for the next phase of the DARPA Robotics Challenge, the first physical challenge. This will be held in December in an outdoor, public event at Homestead-Miami Speedway in Florida.

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. Three hundred seventy-five 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, available across a wide range of devices, enabling everyone on the planet to collaborate and share data and information.

Recent Focus

In recent years, a great many factors (people, devices, bandwidth, policy decisions, and so on) 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, television, publishing, automotive, entertainment, games, 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.

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 how HTML5 will be the cornerstone for this platform. But the full strength of the platform relies on many more technologies, which W3C and its partners are creating, including cascading style sheets (CSS), scalable vector graphics (SVG), web open font format (WOFF), real-time communications (WebRTC), the semantic web stack, extensible markup language (XML), and a variety of application programming interfaces. The platform continues to grow, and the W3C community, in turn, is growing to meet the demand.

The demand is also driving W3C to expand its agenda and the size of its community. W3C launched Community and Business Groups in 2011. In less than two years after launch, nearly 3,000 people participated in more than 120 groups. 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.

The Open Web Platform

The Open Web Platform, a cornerstone of the W3C standardization effort, is a suite of technologies that is transforming business practices, creating new business models, and allowing for greater innovation on the web while reducing product lifecycle costs compared to other systems.

W3C is designing royalty-free technologies that:

- Provide a rich interface feature set including style, interaction, and media
- Enrich mobile apps through APIs for device capabilities and user data
- Integrate data and services (mashups, integration of existing databases and services)
- Run on any device (computer, telephone, television, consumer electronics, automobile, and so on), and support interaction through a variety of input and output modes
- Meet network and communications demands (cross-origin resource sharing, real-time communications)
- Satisfy performance and distribution requirements, enable rapid development and deployment, and facilitate maintenance
- Address diverse social requirements for privacy, security, multilingual content, and accessibility

New Centers and Initiatives

Industry Affiliates Program

Established in 2008, CSAIL's Industry Affiliates Program (IAP) is a membership-based program that serves as a portal for industry at CSAIL. IAP's mission is to enable connections and strengthen collaborations between CSAIL and its industry partners. IAP has 42 members representing a variety of industry sectors and regions, including North America, Asia, and Europe. Key benefits of the program include:

- Keeping abreast of the latest research at CSAIL
- Identifying opportunities to foster and explore new research collaborations
- Engaging with CSAIL students for recruiting

Members of CSAIL-IAP are given access to new technologies and ideas as they move from laboratory to marketplace. Throughout the year, IAP hosts events, seminars, virtual meetings, and activities to help connect member companies with the lab. To facilitate collaboration and communication with industry partners, IAP maintains a members-only website, produces a quarterly newsletter, and publishes a CSAIL student profile book. IAP's flagship event is its annual meeting held in May each year. The event is a two-day conference that showcases CSAIL's research and students and serves as a forum for members to network, connect with, and learn from each other. More than 100 people from 35 member companies attended the 2013 IAP meeting. All activities are designed to help members to connect with CSAIL researchers to create value for their organizations.

bigdata@csail

The goal of the MIT Big Data Initiative at CSAIL, a multi-year effort launched in May 2012, is to identify and develop new technologies needed to solve next-generation data challenges, which will require the ability to scale well beyond what today's computing platforms, algorithms, and methods can provide. Our objective is to enable people to leverage big data by developing tools and platforms that are reusable, scalable, and easy to deploy across multiple application domains.

Our approach includes two important aspects. First, we will work closely with industry and government to provide real-world applications and drive impact. Promoting in-depth interactions between academic researchers, industry, and government is a key goal. Second, we believe the solution to big data is fundamentally multi-disciplinary. Our team includes faculty and researchers across many related technology areas, including algorithms, architecture, data management, machine learning, privacy and security, user interfaces, and visualization, as well as domain experts in finance, medical, smart infrastructure, education and science. The initiative focuses on four broad research themes: computational platforms, scalable algorithms, privacy and security, and applications. In 2013, we kicked off a series of workshops on key challenges in big data: Big Data Integration (April 2013) and Big Data Privacy (June 2013).

MIT Center for Wireless Networks and Mobile Computing

Wireless@MIT is an interdisciplinary center whose goal is to develop the next generation of wireless networks and mobile devices. With its headquarters at CSAIL, the center creates a focal point for wireless research at MIT and addresses some of the most important challenges facing the wireless and mobile computing fields. In particular, the center is innovating in four important areas: improving the use of the wireless spectrum; better networking and better systems for emerging mobile applications (including video/collaboration, transportation, and healthcare); security and privacy; and low-power design.

The work at Wireless@MIT includes contributions from more than 50 MIT faculty members, research staff, and graduate students. A key feature of the center is that it encourages collaborations between MIT faculty in various departments and labs. In particular, the center has faculty members from the Department of Electrical Engineering and Computer Science (EECS), the Department of Mechanical Engineering, and the Sloan School of Management. Its members also come from multiple labs, including CSAIL, the Laboratory for Information and Decision Systems, Microsystems Technology Laboratories, and the Research Laboratory of Electronics.

We launched the center in October 2012 with an event at MIT that included participation from a large number of senior researchers and engineers from member companies. By January 2013, the center had eight members: Amazon.com, Cisco Systems, Intel, MediaTek, Microsoft, STMicroelectronics, Telefonica, and Google. The center launched a high-profile lecture series including a fireside chat with the chairman of the Federal Communications Commission, Julius Genachowski. This attracted a packed audience, was webcast live to our member companies, and inspired numerous questions. The center has been frequently featured in the news, including in the *Boston Globe*, National Public Radio, the *Guardian*, *Discovery* magazine, the *IEEE Spectrum*, and *MIT Technology Review*. Wireless@MIT has been selected by the *Electronic Engineering Times* as one of the 10 electronics visionaries to watch.

Rethinking Artificial Intelligence

Rethinking artificial intelligence (AI) is exploring a number of unconventional approaches to the fundamental goal of developing a computational model of human-like intelligence. The projects in this research initiative examine the question from a number of angles.

Work on *learning to abstract* starts from the premise that abstraction remains a key obstacle in learning. To learn abstractions from observations, we need rich probabilistic representations that support creation of templates and that enable efficient assessment of the predictive value of templates, i.e., the ability to quickly recognize when and where templates may apply. Two key insights provide stepping stones toward these goals: templates must be designed in a way that naturally enables step-wise generalization, and we must develop structural hashing methods that make possible fast hierarchical recognition and tracking of templates.

Our work so far has focused on developing and implementing general ways of performing successive compression by creating and using lattices of building blocks. In particular, we have focused on the efficiency gained by optimizing how the lattice is traversed. We are now testing the approach using cartoon images where useful abstractions—the building blocks—are easier to evaluate. We believe our framework is particularly well suited to learning abstractions across different modalities and hope to demonstrate cross-modal transfer in the context of learning to perform visually guided actions.

Work on *algorithmic encoding* seeks to understand how we can guarantee computational results with unreliable, weakly communicating components. We want to adaptively coax a set of noisy, uncertain components collectively to achieve the task at hand. This is a long-standing problem with major implications across a wide range of fields, including AI, fault-tolerant computing, error-correcting codes, bioengineering, and neuroscience. Unlike previous work in fault tolerance, however, we do not conceive of the task as guaranteeing that every operation succeeds (or that recovery is possible when one does not). Instead, we assume that errors are commonplace and heterogeneous, and seek to redefine how algorithms should be written in this context.

Recent progress in this area includes answering the question of the relation between errors and outcomes. We have derived a theoretical understanding of distributions over outcomes induced from errors. The results highlight the ways in which many local modifications of computations influence global outcomes. Turning this around, we have the means to use noisy distributed systems for modeling distributions over global arrangements. We are currently working on characterizing what computations can and cannot be performed reliably in this setting.

Work in *probabilistic programming* combines Bayesian probability with universal computation. Instead of modeling joint distributions over a set of random variables, probabilistic programs model distributions over the execution histories of programs, including programs that analyze, transform, and write other programs. Probabilistic programs, as the first computationally universal systems for probabilistic modeling, have capacities for abstraction and recursion that can support reflective reasoning, natural language semantics, and learning model structure as probabilistic inference. They can naturally express core commonsense knowledge about physical objects, intentional agents, and their causal interactions, and are starting to be used in cognitive science to model people's intuitive physical and psychological reasoning. They thus have the potential to provide a new approach to AI by bringing the full power of probabilistic inference to bear on classic AI problems that have been largely left out of the statistical AI revolution of the past two decades.

We have seen encouraging early successes from this approach: we have been able to build systems for topic modeling from text, for clustering, reading degraded and adversarially obscured text, and for inferring 3D road models from real-world images. We built all these using probabilistic programs that are 10 to 50 lines long—this is roughly 1/100th the number of lines required in standard baseline approaches—and we built them without any custom inference code. Our current research is focused on hardening and optimizing our systems, developing debugging and profiling tools for probabilistic programs, integrating recent work on massively parallel inference and new relevant compiler techniques, developing new integrated perception and reasoning systems, and broadening and deepening the set of applications in conjunction with our academic and industrial partners.

Our work in *parsing the minimalist program* is motivated by the flood of data available in very different modalities widely used to train programs through machine-learning algorithms, including web-based text (e.g., Google's one trillion n-gram corpus used to bootstrap state-of-the-art machine translation engines), audio recordings collected from thousands of hours of recorded telephone conversations (used to train speech recognition models), and social networking text (used to provide personalization services and build recommendation engines). Much of the emphasis in the language research community has been focused on extracting surface regularities from this data. But surface regularities are insufficient, both empirically and as a research agenda.

Our work attempts to build a bridge between big-data surface regularities and models of deep constraints on representation and structure, thereby taking natural language understanding to a new level unattainable by either approach alone. These deep constraints include constraints on constituency, argument structure, subject-verb agreement, and relative hierarchical ordering between syntactic elements. We are developing ideas and algorithms that can taking advantage of both kinds of information, an approach that we believe will lead to substantially more powerful natural language understanding systems.

entrepreneurship@CSAIL

In fall 2012, CSAIL established the entrepreneurship@CSAIL initiative to help CSAIL research make the transition to start-up-based commercialization, as well as enabling CSAIL students to pursue start-up ideas of their own design. Following a survey and organizational meeting in fall 2012, the subject 6.S078 Entrepreneurship Project was taught in spring 2013 through EECS with 27 registered students and the participation of three local venture capital firms. Fifteen different teams participated in the subject, and self-selected teams made presentations to a panel of venture experts at the end of the term.

Research Highlights

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

Secure Cloud Applications

The web is a very attractive application platform, offering zero-effort application deployment and low-cost hosting through the use of cloud computing. However, the platform consists of disparate technologies whose integration is riddled with security issues. Many recent security attacks rely on data being passed incorrectly across technologies such as structured query language (SQL) injection and cross-site scripting, or on missing access control checks to access data without authorization.

The current web application development methodology makes providing security guarantees a very expensive proposal. Developers are expected to inspect every single line of code to ensure the absence of vulnerabilities. Furthermore, vulnerability inspection requires a very good grasp of the many complex interactions between different technologies employed by a web application, such as the database server, the application server, the server's operating system and file system, an ever-growing spectrum of client-side technologies, and browser-specific vulnerabilities.

Data flow assertions promise to turn security into a feature that can be mixed into an application without the need for cross-cutting changes, and thus is suitable for real-world application development. This is accomplished by marking all the data coming into a system from untrusted sources, such as user input, as tainted, and forbidding tainted data from leaving the system before it is untainted by special methods. For example, strings coming from a hypertext markup language (HTML) form should not be merged into an application's HTML output, or into an SQL query, without being processed by appropriate sanitization functions. Dataflow-based security relies on tracking tainted data, which can be automated, to replace tedious checks spread throughout an application's code base.

We have investigated the application of data flow assertions to the Ruby programming language and to the Ruby on Rails application framework. We aim to stand out from prior work by using Ruby's dynamic nature and meta-programming support instead of relying on interpreter changes or source code transformation, by adapting the policy approach introduced in Resin to the modern development style adopted by Rails, and by reaching an acceptably low performance overhead.

Ruby is a very dynamic language. At any time, the programmer can modify system classes, such as String, by adding or changing methods. Each object has its own hidden super-class (known as the object's eigenclass) that can be used to override methods at the individual object level. Methods can be grouped into modules, which can then be included into classes at runtime. These features allow for a straightforward taint-tracking implementation that does not require interpreter changes. At the same time, modern Ruby interpreters obtain a performance boost by relying on the fact that Ruby programs are mostly static in nature, so using the language's dynamic features has a sharp performance cost. A well-performing taint-tracking implementation requires a deep understanding of the interpreter implementation. We produced a straightforward taint-tracking implementation, and then used the Rubinius interpreter to experiment with different approaches to taint tracking, thereby understanding the interpreter's performance characteristics. This research is under the direction of professor Srini Devadas.

Fast Fourier Transform Algorithm

During the first half of 2012, professors Piotr Indyk and Dina Katabi (together with students Haitham Al-Hassanieh and Eric Price) discovered new, highly efficient algorithms for computing the spectrum of a digitized signal. The algorithms apply to real-world signals that are sparse, i.e., their spectra have few large frequencies. For such signals, the new algorithms are significantly more efficient than the well-known fast Fourier transform algorithm, sometimes by orders of magnitude. This is because their complexity is linear in the number of large coefficients, not the actual signal length. The efficiency of the algorithms is two-fold: the running time is low, and the algorithms sample only a small fraction of the signal.

Over the past year, we have discovered an even more efficient version of the algorithm that further reduces the number of selected samples by an order of magnitude. Further, the algorithms and their variants were shown to help reduce the computation time, power consumption, and device cost for a diverse set of tasks, including magnetic resonance spectroscopy, computational photography, low-power spectrum sensing, and global positioning system locking. We are currently investigating applications to other areas, such as seismic data processing and astronomy. This research is under the direction of professor Piotr Indyk.

Human Robot Collaboration in Assembly Manufacturing

Assistant professor Julie Shah and her research group have been working on designing new algorithms for robot learning, decision making, and control that enable close physical collaboration between human workers and industrial robots. The team applies these technologies to increase the efficiency of aerospace, automotive, and electronics assembly manufacturing. Highlights of the past year's work include the following projects.

Graduate students Matthew Gombolay and Ronald Wilcox, together with Julie Shah, developed a fast task assignment and scheduling algorithm that enables multiple robots to quickly reorganize their work assignments in response to changing guidance from a human coworker or supervisor. The technique provides strong guarantees that the schedule meets synchronization, timing, and spatial-proximity constraints, which is necessary to maintain safety when humans and industrial robots maneuver in a shared workspace. The novel algorithms produce near-optimal solutions for up to 10 robots and 500 tasks in less than 10 seconds on average, which is an order of magnitude increase in problem size and more than two orders of magnitude decrease in computation time, compared with prior performance. MIT has filed two patent applications on this work. The team is currently working to apply the algorithm to robot tasking problems in a factory for aerospace manufacturing.

Other work, performed in collaboration with graduate student Stefanos Nikolaidis, focused on developing new algorithms for training humanoid industrial robots to work in teams side-by-side with people. The work draws insight from effective techniques for human team training to enable a robot to adapt and customize a high-level task plan through practice working with the human partner. Stefanos Nikolaidis and Julie Shah redesigned a standard interactive robot learning algorithm to better emulate a "gold standard" human team training technique called cross-training, in which the human and robot learn team fluency by switching roles. Cross-training is empirically documented to improve human team coordination and is used widely in military, aviation, manufacturing, and disaster response teams. Experiments with human-robot teams demonstrated that cross-training produced statistically significant improvements in human-robot team performance, including a 41 % reduction in the human's idle time. This work was featured in MIT News, the New York Times, and other media outlets. This research is under the direction of assistant professor Julie Shah.

Interpreting the Molecular Basis of Human Disease

The completion of the mapping of the human genome at the turn of the century opened up a decade of biological discovery aimed at understanding how genetic variants lead to inherited human disease. To that effect, the genotyping of increasingly large cohorts of patients and healthy individuals has resulted in genome-wide association studies that report the frequency with which both ancestrally inherited and recently occurred polymorphisms in the human population are associated with a wide range of diseases and other traits. The challenge to translate these associations into precise hypotheses at the molecular level about the specific molecular mechanisms leading to disease onset is one of the greatest challenges in modern biology and medicine. It involves not only understanding the function of any nucleotide in the three billion letters of DNA in the human genome, but also predicting the effect of any change. Our work has focused on large-scale data integration of comparative genomics, functional genomics, and population genetics datasets to understand the circuitry of human disease.

This past year, we published the integrative analysis of the Encyclopedia of DNA Elements project, the culmination of five years' research conducted through a large international collaboration. The project generated the most comprehensive annotation of the human genome yet known, providing a unique map for interpreting genetic information from disease association datasets. In particular, only 1% of the genome is encoding proteins while the remaining 99% remains largely unannotated. Our group provided a comprehensive annotation of regulatory regions in multiple human cell types, resulting in descriptions of millions of previously uncharacterized regions that can be mined to interpret the putative functions of almost any region of the human genome. We analyzed the DNA sequence motifs associated with regions of regulator binding, giving us unprecedented insight into the lexicon of human regulatory regions that govern tissue-specific gene expression patterns. We also discovered thousands of regulatory regions whose activity differs between maternal and paternal chromosomes within the same cell, uncovering regions of imprinting (parent-of-origin effect) and regions where sequence variants between maternal and paternal chromosomes result in different sequence motifs. This study was published in *Nature* in September 2012.

We conducted the first investigation of evolutionary constraint within the human lineage associated with annotated regulatory regions in multiple cell types specifically tuned outside regions of mammalian conservation. We found that, in addition to the approximately 5% of the human genome that lies within mammalian-conserved regions, an additional 4%–11% shows population genetic signatures of constraint within the human lineage. These signatures are based on the allele frequencies of different classes of regulatory regions aggregated across the whole genome and over multiple individuals. This study answered a long-standing question about the discrepancy between genome-wide functional genomics experiments, which show abundant biochemical activity over large fractions of the genome, and comparative genomics studies, which show only a small fraction of the genome under constraint. The results suggest that regulatory regions of the human genome experience rapid turnover, consistent with dramatic rewiring. This study was published in *Science* in September 2012.

We also developed a new method for systematic testing of the thousands of regulatory genomics predictions that our algorithms make about the effect of individual variants, to circumvent the traditional low-throughput approaches for testing variants. We used this method to dissect the roles of regulatory motifs and individual single-nucleotide changes in 2,000 human enhancers, conducting a total of 50,000 experiments, and succeeded in demonstrating the importance of regulatory motifs and their context, evolutionary conservation as a metric for identifying important regulatory motifs, and the epigenomic features most predictive of functional distal regulatory elements known as enhancers. This study was published in Genome Research in March 2013.

Finally, given the very large number of interconnected components in our regulatory models, we also developed new algorithms for studying regulatory networks and elucidating the most meaningful relationships among them. In particular, as most gene networks rely on correlated patterns of activity to distinguish interconnected elements, the field of network studies is plagued with an abundance of spurious interconnections due to indirect effects between nodes. We took a spectral matrix analysis approach to this problem, reasoning that the effects of indirect edges can be modeled as successive matrix multiplication operations of an unobserved direct-effect matrix and successive summing of these second-order, third-order, etc. effects in an infinite series of diminishing returns (“network convolution”). By decomposing the observed adjacency matrix into its eigenvalues and eigenvectors, we could reverse the effects of network convolution for each eigenvalue separately, and recombine these into the original direct matrix in a new algorithm (“network deconvolution”). We applied this method outside the field of gene networks and found that it is also applicable to protein folding, by distinguishing direct amino-acid interactions from residue correlations across multiple sequence alignments, and to social networks, by distinguishing strong collaborations from coauthorship networks based on the full structure of the network. This study was published in *Nature Biotechnology* in July 2013. This research is under the direction of associate professor Manolis Kellis.

Practical Computation Over Encrypted Data

Disclosures of confidential data are commonplace. This often happens as a result of adversaries breaking into a server and gaining access to all of the confidential information stored there. A promising approach to protect confidential data from such attacks is to encrypt it: even if unauthorized persons somehow gain access to encrypted data, they will not be able to decrypt it, as long as they do not have the decryption key. The main problem with this approach is that the server often requires access to this key to decrypt the data for processing and as a result someone who breaks into the server can gain access to the decryption key as well.

In our research, we developed several practical techniques that enable applications to process encrypted data—in particular to execute queries over encrypted databases—without having to decrypt the data at the server. Using these techniques, servers that store and process confidential data no longer need access to the decryption key and the data remains confidential even if an third party compromises a server.

One fundamental operation that database queries may perform is to compare the order of items; this is a building block for sorting, range queries, and many other computations. An order-preserving encryption scheme (that is, a scheme where the encryption of A is greater than the encryption of B if and only if A is greater than B) allows computing the order of two items by simply comparing their encryptions. We developed the first such scheme that provides ideal security; that is, the encrypted values reveal only the order of the original values and nothing more. A key insight was to relax the strict requirements traditionally imposed on an encryption scheme, such as being immutable and stateless.

In addition to order-preserving encryption, we use a range of different encryption schemes to efficiently perform simple operations on encrypted data, such as addition, keyword search, and equality comparison. To execute more complex queries not supported by any one of these schemes we developed a technique to split a complex query into both simple pieces that can run over encrypted data and more complex parts that have to run on a trusted client machine with access to the decryption key. By using an optimizing query planner, this system is able to minimize the cost of split query execution and achieve modest overheads, on the order of 24% to meet the industry-standard TPC-H benchmark. This research is a collaboration of PhD students Raluca Ada Popa and Stephen Tu, and professors Frans Kaashoek, Sam Madden, and Nickolai Zeldovich.

Laboratory Sponsored Activities

CSAIL Outreach

Imara

The overall goal of Imara is to find and implement long-term, sustainable solutions to make educational technology and resources available to domestic and international communities. In concert with partners both inside and outside of the Institute, we hope our work will help bridge the digital divide—the perceived gap between those who have access to the latest information technologies and those who do not.

CommuniTech

CommuniTech is a domestic outreach initiative that attempts to heal the digital divide in the local community. The program provides economically disadvantaged adults with the tools they need to gain access to valuable information they can use to better their lives and the lives of their families.

CommuniTech focuses on teaching basic computer technology skills and providing access to computer hardware. Families Accessing Computer Technology (FACT) is a sixweek course, taught by MIT student volunteers, designed to teach basic computer skills to underprivileged adults. Clients gain basic computer proficiency, marketability to prospective employers, and a stronger connection to the world at large.

Clients who complete the class receive computers from FACT's partner program, the CSAIL Used Computer Factory (UCF). UCF refurbishes donated computers by installing new operating systems and productivity software. UCF also plays a small part in "greening" MIT by recycling unwanted equipment that might otherwise end up in landfills.

Lacotec Laare, Kenya

The Laare Community Technology Centre (LCTC) in Kenya was founded by Eric Mibuari ('06) in 2005 with the assistance of MIT's Public Service Center. The aim of this community center is to increase general computer awareness and literacy in the Laare community by providing cheap, local, and accessible training in the use of computers.

LCTC particularly targets young people with high school educations, seeking to equip them with basic computing skills that they can apply for personal use, in industry, and in their education. So far, LCTC has offered various levels of training to more than 500 students. LCTC takes particular note of the economic difficulties of its potential trainees and strives to charge the lowest feasible fees. Recently, it has expanded to include a primary-school IT educational initiative for elementary school students.

Middle East Education through Technology

CSAIL's support of the Middle East Education through Technology (MEET) program has continued over the past year. MEET is an innovative educational initiative aimed at creating a common professional language between young Israelis and Palestinians. MEET enables its participants to acquire advanced technological and leadership tools while empowering them to create positive social change within their own communities. Many MIT students volunteer to teach MEET summer courses at the Hebrew University in Jerusalem. CSAIL continues to host http://meet.csail.mit.edu/ and to provide technical support to the MEET program.

TEK

Several CSAIL members have been supporting the Time Equals Knowledge (TEK) project. The project empowers low-connectivity communities by providing a full internet experience using email as the transport mechanism.

Although the internet has revolutionized information delivery for most of us,, accessing online resources remains an economic and technological challenge for many communities in the developing world. High charges for telephone connections and for internet access can quickly grow unaffordable, and low-bandwidth connections limit the amount of material that can be viewed per session. Because phone lines are limited, it is often difficult to time-share between internet and voice. Furthermore, unreliable network and power infrastructures can sometimes block internet access altogether.

Compared to direct web access, email can be much cheaper, more reliable, and more convenient in developing areas. The TEK client operates as a proxy on a user's machine, enabling users to browse downloaded pages using a standard web browser. New searches are automatically encoded as emails and sent to the TEK server, which queries the web and returns the contents of resulting pages via email. TEK is free software distributed under the GNU lesser general public license.

OpenCourseWare Outreach Initiative

Adnan Esmail, a mechanical engineering graduate student, has combined MIT's OpenCourseWare (OCW) initiative with CSAIL's support to bring the institute's educational resources to the Indian subcontinent. The OCW outreach initiative consists of mirror sites that make MIT course content information available to those in nations with poor bandwidth infrastructure.

The program has been realized in partnership with Aligarh Muslim University in Aligarh, India, and Lahore University of Management Sciences, in Lahore, the capital of Pakistan's Punjab province. It runs on hard drives that have been generously donated by Maxtor for OCW proliferation. This expansion of access gives talented students who would otherwise be without the technological resources they need a chance to reach their full educational potential.

Dertouzos Distinguished Lecture Series

Three speakers gave presentations during the 2012–2013 Dertouzos Distinguished Lecture Series. They were:

- Andrew Howard, Space Exploration Technologies, "Enter the Dragon: The SpaceX COTS Missions"
- Rodney Brooks, Rethink Robotics, "A New Class of Industrial Robot"
- David Culler, University of California, Berkeley, "Software Defined Buildings—A Computer Systems Approach to Making the Built Environment Better and More Sustainable"

Organizational Changes

In May 2012 professor Daniela Rus was named 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. Previously, Professor Rus served as associate director of CSAIL from 2008 to 2011 and has been codirector of CSAIL's Center for Robotics since 2005. She also directs CSAIL's Distributed Robotics Laboratory. Daniela Rus is the first woman to serve as director of CSAIL.

Professor Rus succeeded professor Anant Agarwal, who became the president of edX, a nonprofit online learning initiative founded by MIT and Harvard University. The initial platform for this online initiative was developed at CSAIL.

CSAIL's leadership team includes two associate directors and an executive cabinet. Persons holding these positions are appointed by the laboratory's director. Professor Randall Davis and professor Robert Miller became associate directors for the 2012–2013 academic year and assist the director with her duties.

The CSAIL executive cabinet meets twice per month to review and advise the director on policy, processes, and activities within the laboratory. Members named to the executive cabinet include Regina Barzilay, Jack Costanza, Randall Davis, Srini Devadas, John W. Fisher, David Gifford, John Guttag, Daniel Jackson, Robert Miller, Martin Rinard, Ronitt Rubinfeld, Christopher Terman, Karen Shirer, and Victor Zue.

Srini Devadas took over the role of "space czar," overseeing the space committee and managing the allocation of space within CSAIL. The space committee implements improvements to the facilities that will increase the quality of the environment for the laboratory's faculty, staff, and students. John Guttag, the former space czar, became a member of the space committee to assist with the transition to new leadership. The committee also includes the two assistant directors, Jack Costanza and Karen Shirer, and administrative assistant Sonya Kovacic.

Jack Costanza continued as the assistant director for infrastructure, overseeing information technology infrastructure and user support, building operations, and communications. Karen Shirer continued in her role as assistant director for administration, overseeing finance and human resources.

In May 2012, CSAIL launch a new initiative, bigdata@CSAIL, to tackle the challenges of the burgeoning field known as "big data." The imitative strives to address the issues that can arise when data collections are too big, growing too fast, or too complex for existing information technology systems to handle. In December 2012, Elizabeth Bruce, CSAIL's director of industry partnerships, was named the first executive director of this initiative.

In May 2013, Lori Glover was named the new director of the CSAIL Industry Affiliates Program, a corporate membership program that offers companies the opportunity to access CSAIL's faculty and students through annual conferences, recruiting events, and on-site visits.

Awards and Honors

Our faculty and staff won many awards this year, including the following:

> Anant Agarwal: National Academy of Engineering, Member; *Forbes* Magazine, Top 15 Education Innovators; American Academy of Arts and Sciences, Fellow
>
> Saman Amarasinghe: Association for Computing Machinery (ACM)/Institute of Electrical and Electronics Engineers (IEEE), Most Influential Paper from the International Symposium on Code Generation and Optimization
>
> Bonnie Berger: International Society for Computational Biology, Fellow
>
> Tim Berners-Lee: Royal Academy of Engineering, Queen Elizabeth Prize for Engineering

Britton “Bryt” Bradley: MIT EECS, Richard J. Caloggero Award

Adam Chlipala: National Science Foundation, Career Award

David Clark: ACM Special Interest Group on Data Communication (SIGCOMM) Test of Time Award

Erik Demaine: Science Atlantic, Outstanding Student, Hall of Fame; European Association for Theoretical Computer Science, Presburger Award; John Simon Guggenheim Memorial Foundation, Guggenheim Fellow

Jack Dennis: IEEE, John von Neumann Medal

William Freeman: IEEE, FG 2013 Test of Time Award

Shafi Goldwasser: The Simons Foundation, Simons Investigator; ACM A.M. Turing Award

Polina Golland: MIT, Jamieson Prize for Excellence in Teaching

Piotr Indyk: MIT EECS, Faculty Research Innovation Fellowship; ACM, Paris Kanellakis Theory and Practice Award; The Simons Foundation, Simons Investigator

Dina Katabi: ACM, Grace Murray Hopper Award and Doctoral Dissertation Award

John Leonard: IEEE International Conference on Robotics and Automation, Best Paper Finalist

Barbara Liskov: National Academy of Inventors, Charter Member; Technical University of Catalonia, honorary doctorate; ACM, Special Interest Group on Operating Systems Hall of Fame Award; Massachusetts Academy of Science, Fellow

Andrew Lo: Carnegie Mellon University, Nash Distinguished Lecture; American Academy of Arts and Sciences, Fellow

Nancy Lynch: Carnegie Mellon University Computer Science Department, Distinguished Lecture; Logic in Computer Science Conference, keynote address; ACM, Athena Lecturer Award

Samuel Madden: CSAIL, Best Paper Award—Conference on Innovative Database Research; Conference on Innovative Data Systems Research, Best Paper Award; ACM, EuroSys Conference Best Paper Award

Wojciech Matusik: DARPA, Young Faculty Award

Silvio Micali: ACM, A.M. Turing Award

Robert Miller: MIT, MacVicar Faculty Fellow

Una May O’Reilly: Evolutionary Design and Optimization Group, Best Paper

Ronald L. Rivest: National Cyber Security Hall of Fame, National Cyber Security Hall of Fame Award

Daniela Rus: 2012 Robot Film Festival, Best Robot Actor Award; IEEE International Conference on Biomedical Robotics and Biomechatronics, Best Paper Finalist; IEEE International Conference on Robotics and Automation, Best Student Paper Finalist; Massachusetts Women’s Forum, Elected

Armando Solar-Lezama: Conference on Innovative Data Systems Research, Best Paper Award

Karen Sollins: American Association for the Advancement of Science, Fellow; ACM SIGCOMM, Test of Time Award

Madhu Sudan: American Mathematical Society, Fellow

Peter Szolovits: American College of Medical Informatics, Morris F. Collen Award of Excellence; American Medical Informatics Association Annual Symposium, Best Student Paper

Daniel Weitzner: *Newsweek/Daily Beast* Digital Power Index; International Association of Privacy Professionals, Privacy Leadership Award

Victor Zue: IEEE, James L. Flanagan Speech and Audio Processing Award; Okawa Foundation for Information and Telecommunications, The Okawa Prize

Key Statistics for Academic Year 2013

Faculty: 105 (19% women)

Research staff: 29 (16% women)

Administration, technical, and support staff: 71 (66% women)

Postdocs: 63 (11% women)

Visitors: 111 (14% women)

Paid Undergraduate Research Opportunities Program participants: 105 (37% women)

Master of engineering students: 44 (36 graduate students: 344 (19% women)

Daniela Rus
Director
Professor of Computer Science and Engineering