Computer Science and Artificial Intelligence Laboratory

The Computer Science and Artificial Intelligence Laboratory (CSAIL) pioneers research in computing that improves the way 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.

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 that confront us.

CSAIL researchers and CSAIL operations adapted well to work during a pandemic. The Covid-19 pandemic has brought many of the inequities of our economic systems front and center; it has also highlighted some bright opportunities for building a better world. One of those bright spots is the rapid development of new technology, where CSAIL research has had the opportunity to contribute. For example, our team developed and helped deploy an Ultraviolet-C disinfecting robot in the Johnson testing facility. Our researchers are actively developing machine-learning (ML) systems and computational analyses that are helping researchers understand the virus, predict how to provoke an immune response, choose the elements of potential vaccines, make sense of experimental data, and even track the virus' genetic mutations over time.

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

CSAIL leadership spearheaded several other efforts that benefited the broader MIT community:

Achievements

New Programs

In May 2021, Mentored Opportunities in Research launched the diversity postdoctoral program, including the Interval collaboration for computer-focused startups as well as AI accelerator training for the Department of Defense and national security leaders.

Communities of Research

Launched on January 1, 2020, nine CSAIL Communities of Research (CoR)—a new system of self-governance and identity—support the CSAIL community in social and intellectual impact through administrative support for CoR heads and by developing values for the community. In AY2021, CoR additionally assisted in the search for a director of communications and for several fiscal positions.

Campus Collaborations

CSAIL collaborated with the Institute for Medical Engineering and Science to host the AI in Medicine Trans-Atlantic Symposium.

CSAIL Constitution

Reviewing and codifying operations for increase transparency and accommodations of changes to work life after the pandemic include decision-making and codification of CSAIL ES, completing the Space Management and Proposal submission, and more.

Developing programs—such as the COVID-19 Response, CSAIL Graduate Student & Postdoc Council, and CSAIL Research office hours—support the community emotionally and financially while inspiring research.

New Lab-wide Research Initiatives

CSAIL launched funded initiatives to address basic research challenges in AI, resilience, and unsupervised learning, such as Defense Science and Technology Agency, MLA@CSAIL, and more.

CSAIL Growth

In the past ten years, we have had a 75% growth in research volume. With a total combined research volume (primary and secondary funds) of $84,136,682 for FY2021, CSAIL is the Institute's largest interdisciplinary research laboratory and continues to have the highest research volume amongst MIT Interdisciplinary Departmental Laboratories.

CSAIL manages approximately 619 active research awards and over 124 PIs with appointments across 11 MIT departments. Through academic year 2021, CSAIL had 494 graduate students with research associate appointments and 186 UROP students.

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 following:

- Air Force Research Laboratory and the Air Force Office of Scientific Research
- Army Research Office
- Defense Advanced Research Project Agency
- Department of Defense Research and Engineering

- Food and Drug Administration
- Department of Education
- Department of Energy
- Intelligence Advanced Research Projects Activity, National Institutes of Health
- National Institute of Justice
- National Science Foundation
- US Navy (including the Office of Naval Research, and Naval Air Systems Command)
- Space and Naval Warfare Systems Center

US and international nonfederal sponsors include the following:

- Accenture
- Advanced Technology Laboratories
- Aptima
- BAE Systems
- BBN Technologies
- Boeing
- BMW of North America
- Delta Electronics Foundation
- DSTA
- Ford Motor Company
- Foxconn Technology Group
- Honda R&D
- IBM
- Intel Corporation
- JD.com
- Jaguar Land Rover Limited
- Lockheed Martin
- Microelectronics Advanced Research Corporation
- Mitsubishi Electric Corporation

- National ICT Australia Limited
- Nissan Motor Company Limited
- Nippon Electric Company
- Nippon Telegraph and Telephone Corporation
- Northrop Grumman Corporation
- Omron
- Ping An Technology
- Pfizer Inc.
- Quanta Computer
- Rakuten
- Raytheon
- Steelcase Incorporated
- Systems & Technology Research
- Takeda
- Samsung Electronics
- Siemens
- Suzhou Industrial Park
- Toyota Research Institute
- Wistron Corporation

Other organizations sponsoring research include:

- Aarhus University
- Battelle Memorial Institute
- DSO National Laboratories
- Epoch Foundation
- The Hong Kong University of Science & Technology
- Industrial Technology Research Institute
- Nanyang Technical University
- Singapore-MIT Alliance

Research Projects

Within CSAIL we have 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 the following:

Defense Science and Technology Agency

In July 2020, CSAIL launched a research collaboration with the Defense Science and Technology Agency in Singapore. The program includes the following projects which were initiated in January 2021:

SYNTHBOX: Establishing Real-World Model Robustness and Explainability Using Synthetic Environments

Led by Professor Aleksander Madry, the ultimate goal of this project is to leverage recent advances in realistic graphics rendering and the expertise of the principal investigator (PI) in data modeling and inference to build SynthBox—a versatile framework for creation, manipulation and semi-automated analysis of synthetic environments with the goal to establishing robustness of real-world models. Specifically, this framework will enable one to get a complete and semi-automated control over the synthesized scenes so as to use it for fine-grained analysis, and remedying, of the vulnerability of vision models train on the real data to a rich set of perturbations and data corruptions/distribution shifts. The ultimate ambition is to devise a toolkit for real-world deployment of robust models that can seamlessly accommodate a wide range of sophisticated robustness requirements (e.g., robustness to repainting of the object, change of lighting, realistic obstructions) and, crucially, has minimal real-data acquisition and manipulation requirements. This framework will also support evaluation, and development, of explainability methods by providing the user with a wide-range of counterfactual scene transformations.

Next Generation Natural Language Processing Technologies for Low Resource Tasks

Led by Regina Barzilay, Delta Electronics Professor, and Tommi Jaakkola, Thomas Siebel Professor in Electrical Engineering Computer Science and Institute for Data, Systems, and Society, the successes of natural language technologies have primarily come from few annotation-rich tasks such as machine translation across select languages. Such supervised learning tasks are well-suited for deep learning architectures similar to object recognition in computer vision.

However, most languages in the world are not richly annotated, and most tasks—including style transfer, fact verification, and many others—are held back as the associated tasks are diverse and annotations scarce or challenging to come by. The lack of direct supervision often results in inaccurate, indefensible, and brittle outputs, unsuitable for driving downstream applications. This project directly addresses these challenges, developing new approaches to effective style transfer and fact verification. The researchers' underlying learning technology builds on and further develops two key pillars: recent advances in invariant prediction for effective extrapolation with scarce annotations as well as explicit data augmentation by textual rewriting. The researchers' methods also contribute directly to interpretable modeling, though this is not the primary focus of the project.

Improving Situational Awareness for Collaborative Human–Machine First Responder Teams

Led by Professor Nicholas Roy, when responding to emergencies in urban environments, an essential need is achieving situational awareness. To address this need, this project will develop a team of autonomous air and ground vehicles designed to arrive at the scene of an emergency, build a map of the scene to provide a situation report to the first responders in advance, and search for people and entities of interest.

Developing a system that can execute such a mission requires solving two main technical challenges. The first challenge is how to fuse data from multiple heterogeneous vehicles to build a coherent model of an emergency scene.

The second challenge is how to plan the paths of the vehicles to get the best situational awareness in terms of locating objects and people as well as ensuring an accurate model for the arriving first responders. The ultimate deliverable of this project is a multi-agent system that can autonomously provide an accurate and complete map of an urban emergency or natural disaster scene.

Trustworthy, Deployable 3D Scene Perception via Neuro-symbolic Probabilistic Programs

To be deployable in the real-world, 3D scene perception systems need to generalize across environments and sensor configurations and adapt to scene and environment changes without costly retraining or fine-tuning. They also need to track objects consistently across long occlusions and use prior knowledge about object motion to constrain scene interpretations while filling in the blanks with physically possible poses. To be trustworthy and explainable, they additionally need to be able to flag input data that appears to depart from their underlying modeling assumptions and to report uncertainty over 3D scene composition, structure, and pose. Together, these capabilities would allow 3D scene perception to be safely deployed in real-world environments and used as the first stage of autonomous systems that need to make robust plans. This project is led by Vikash Mansinghka, principal research scientist, Joshua Tenenbaum, professor of Cognitive Science and Computation, and Antonio Torralba, Thomas and Gerd Perkins Professor and faculty head of Artificial Intelligence and Decision-Making (AI+D).

This project aims to develop and test a domain-general open-source solution to this fundamental trustworthy AI problem, building on the team's breakthroughs in probabilistic programming and in real-time neural Monte Carlo inference for symbolic generative models.

Computationally Supported Roleplaying for Social Perspective Taking

Led by Professor of Digital Media & AI+D Fox Harrell, many issues that stoke conflict—such as bias, discrimination, and misinformation about social groups—are rooted in a failure to understand the perspectives of others. Yet with reflection, it is possible for people's perspectives to change. This project seeks to develop a transformative approach to perspective taking and perspective change using computationally supported roleplaying systems (including games, virtual reality, and artificial reality).

Such roleplaying systems are now pervasive and have been shown to be powerful means of sparking conceptual and behavioral change for users. Drawing on both computer science and social science approaches, this project will result in a framework for designing systems to support users in perspective change through reflection and to model social phenomena such that users measurably better understand the perspectives of others with different social identities (e.g., differences of nationality, ethnicity, gender, values), in addition to a prototype application implementing the researchers' approach.

Analytics-guided Communications to Counteract Filter Bubbles and Echo Chambers

Led by Professor Deb Roy, social media technologies that promised to open our worlds have instead algorithmically driven us into cocoons of homogeneity, exacerbating socioeconomic divides, limiting our exposure to different perspectives, and curbing our opportunities to learn from others who don't necessarily look, think, or live like us.

With their practices and technologies optimized for AI-driven engagement, current news and social media business models are fragmenting us into isolated tribes that are hostile toward outside views and ripe for the spread of misinformation, polarization, hateful discourse, and even violence. The loudest, most extreme voices now dominate the public sphere, stifling constructive communication across divides.

The practical consequences of this situation are clear and dire. Information that should reach all and be understood by all—such as the role of wearing face coverings to slow the spread of Covid-19, or the individual right to vote—are not reaching all of us and are becoming compromised by the spread of misinformation. This project uses two interrelated research threads aimed at counteracting these problems: first, develop language models that predict how "bridging" versus "polarizing" a message will be, enabling communicators to modify word choice to reduce polarization and improve cross-group reach and (potentially) shared understanding. Second, develop methods to monitor the reach of communication campaigns, detect underexposed groups, and find new messengers/channels to reach underexposed groups to maximize message reach across divides.

Decentralized Learning with Diverse Data

This project aims to develop a framework for decentralized learning with multiple agents with different objectives and non-independent and identically distributed data. The goal is to take a holistic approach providing a systematic understanding of the fundamental tradeoffs of pooling and adaptation, developing formulations that can provide efficient and fair performance across agents with diverse objectives, and designing sample, iteration, and communication efficient algorithms. This ambitious agenda will bring tools from machine learning, optimization, control, statistics, statistical physics and game theory.

The researchers will complement the strong theoretical emphasis of their work with a concrete robotics testbed involving dexterous manipulation with diverse objects. Like autonomous vehicles, robotic manipulation exposes the core challenges of learning for perception and control that must be coordinated across a fleet. Obtaining diverse real-world data for vehicles is quite challenging; it requires operating in or simulating a diversity of natural environments. Obtaining diverse data for manipulation is relatively easy and requires only spending a few dollars at the local grocery store. PI Tedrake's lab has many physical robots for testing and has also developed a high-fidelity simulation capability in which we can explicitly study distributions over task and environment parameters (e.g., lighting conditions, object geometry and material properties) and control the distribution shift across time for a single agent and across the fleet. This project is led by Konstantinos Daskalakis, professor in Electrical Engineering and Computer Science (EECS), Asuman E. Ozdaglar, head of EECS and deputy dean of Academics, MIT Schwarzman College of Computing, and Russell L. Tedrake, Toyota Professor and professor of Computer Science and Engineering, Aeronautics and Astronautics (AeroAstro), and Mechanical Engineering (MechE).

Data-driven Optimization under Categorical Uncertainty and Applications to Smart City Operations

Led by Associate Professor Alexandre Jacquillat, this project tackles two questions. From a practical standpoint, it proposes new AI tools to manage the cyber-physical infrastructure in smart cities by jointly optimizing the allocation of information and communication technologies resources (e.g., mobile bandwidth, radio access technologies) and physical resources (e.g., vehicles, personnel) in support of mission-critical operations. From a methodological standpoint, it develops new models and algorithms toward data-driven decision-making under categorical uncertainty, by combining classification outputs from machine learning into optimization methods from mathematical programming.

Provably Robust Reinforcement Learning

Led by Norbert Wiener Professor of Mathematics Ankur Moitra, this project could lead to practical and provably robust algorithms for learning nearly optimal policies in contextual bandits and related problems. These algorithms would be able to tolerate a constant fraction of the rounds being adversarially corrupted, whereas existing algorithms make the much more stringent assumption that the corruptions in each round are bounded. This makes them unsuitable in settings where our machine learning systems are interacting with a large group of agents, some of which are malicious.

Moreover, right now, algorithms that can tolerate adversarial corruptions in reinforcement learning settings are only known in low dimensional settings. For example, Lykouris, Mirrokni, and Paes Leme gave robust algorithms for multiarmed bandits. However, they work based on building robust confidence intervals for each arm separately. In contrast, contextual bandits are truly high dimensional in the sense that we need a function on a high dimensional input that can robustly predict the expected reward for each action. The project will take a major theoretical step of marrying recent progress in robust high-dimensional learning with dynamic online settings with feedback.

Building Dependable Autonomous Systems Through Learning Certified Decisions and Control

Led by Assistant Professor Chuchu Fan a fundamental hurdle in the widespread deployment of modern AI techniques including deep learning and reinforcement learning (RL) on safety-critical autonomous systems is the lack of formal guarantee on safety, robustness, and performance.

Mathematical certificates, which widely exist in formal methods and control theory, serve as proofs that the desired properties of the system, including safety, robustness, and performance, are provably satisfied when the corresponding decision-making, planning, and control components are run in closed-loop composition with the system. The PI's research has created a new framework that enables the co-learning of the planner and control simultaneously with certificates, so the learned planner and control are verifiable correct by construction. This project aims to advance the fundamental science of understanding when the planner and control and certificate can be learned together, what kind of autonomous systems can be certified, how robust the certified systems are, how to transfer the learned results among different systems and environment configurations, and how to extend to large-scale multi-agent systems.

Online Learning and Decision Making under Uncertainty in Complex Environments

Led by Dugald C. Jackson Professor Patrick Jaillet, the main focus of this research is on the fundamental aspects of human sequential decision-making under uncertainty, when information about the past and current environment is accessible through heterogeneous datasets, possibly noisy and/or adversarial in nature, and when the future environment is not perfectly known in advance and may adapt to past decisions.

More specifically, this project will investigate the interplay of three main phenomena: heterogeneous and dynamically generated input data from various sources, including noisy and/or adversarial ones; need for online and real-time decision making; and possibility of online sequential learning about the underlying uncertainty.

This research project will address the following fundamental questions when facing a problem exhibiting one or several of these phenomena: how to properly model and quantify the degree of uncertainty in such a problem, and its impact on the ability to solve it; how to simultaneously optimize and learn in such an environment; and how to design provably good algorithms for helping human decision making.

New Representations for Vision

Led by Thomas and Gerd Perkins Professor of Electrical Engineering William Freeman and Professor Tenenbaum, this project will develop neural network representations for images that are better suited to the requirements for image representations in vision and graphics. These will represent a 3D world efficiently, capturing richness. The representation will allow for efficient rendering to images and efficient inference from images to 3D. Such representations don't exist now. The innovations leading to these representations will be: to augment SIREN networks with a wavelet architecture and to utilize that for a Scene Representation Network, adding constraints to the networks to allow learning a signed distance function and adding uncertainty to the implicit scene representations.

Air Force AI Accelerator

Launched in 2019, the United States Air Force (USAF)-MIT Force AI Accelerator brings together the expertise and resources of MIT and the USAF to perform fundamental research aimed at enabling rapid prototyping, scaling, and application of AI algorithms and systems. Ten projects that encompass a total of 15 research workstreams engage more than 140 faculty, researchers, and students affiliated with more than 20 different organizational units across MIT Campus and MIT Lincoln Laboratory. All project teams involve Air Force personnel, who are embedded in the research teams and also serve as liaisons between AI Accelerator projects and Department of Defense stakeholders. These projects, which began in January 2020, advance AI research in a broad range of areas, including weather modeling and visualization, optimization of training schedules, and enhancement of autonomy for augmenting and amplifying human decision-making. The research activities of the AI Accelerator have been successfully expanding, including seed research projects in collaboration with the Naval Postgraduate School and an AI Education Research project that started in January 2021.

The projects and workstreams with their respective MIT campus PIs, co-principal investigators (Co-PI), co-investigators (Co-Is), and Lincoln Laboratory leads are as follows:

- Guardian Autonomy for Safe Decision Making
    - PI: Professor Daniela Rus (EECS)
    - Lincoln Laboratory Leads: Ross Allen and Ho Chit Siu
- Fast AI
    - PI: Professor Charles E. Leiserson (EECS)
    - Lincoln Laboratory Lead: Vijay Gadepally
- ML-enhanced Data Collection, Integration, and Outlier Detection
    - PI: Professor Tim Kraska (EECS)
    - Co-PIs: Professor Samuel Madden (EECS), Professor Michael Stonebraker (EECS), and Professor Manya Ghobadi (EECS)
    - Lincoln Laboratory Lead: Benjamin Price
- Transferring Multi-Robot Learning through Virtual and Augmented Reality for Rapid Disaster Response
    - PI: Professor Sertac Karaman (AeroAstro)
    - Co-PI: Professor Luca Carlone (AeroAstro)
    - Lincoln Laboratory Lead: Daniel Griffith (Chemical Engineering)
- Conversational Interaction for Unstructured Information Access
    - PI: James Glass (CSAIL)
    - Co-PI: Boris Katz (CSAIL)
    - Lincoln Laboratory Lead: Charlie Dagli

- AI for Personalized Foreign Language Education
    - PI: Professor Shigeru Miyagawa (Linguistics and Philosophy)
    - Co-PI: Professor Emma Teng (History Section)
    - Lincoln Laboratory Lead: Douglas Jones
- Multimodal Vision for Synthetic Aperture Radar
    - PI: Professor Phillip Isola (EECS)
    - Co-PIs: Professor Taylor Perron (Earth, Atmospheric, and Planetary Sciences [EAPS]), Professor William Freeman (EECS)
    - Lincoln Laboratory Lead: Miriam Cha
- AI-Assisted Optimization of Training Schedules
    - PI: Professor Hamsa Balakrishnan (AeroAstro)
    - Lincoln Laboratory Lead: Michael Snyder
- The Earth Intelligence Engine
    - PI: Professor Dava Newman (Media Lab)
    - Co-PIs: Christopher Hill (EAPS), Stephanie Dutkiewicz (Center for Global Change Science)
    - Lincoln Laboratory Lead: Mark Veillette
- Continual and Few-Shot Learning: Transferring Knowledge to New Low Resource Domains and Tasks
    - PI: Professor Pulkit Agrawal (EECS)
    - Co-PIs: Professor Regina Barzilay (EECS), Professor Marin Soljacic (Physics)
    - Lincoln Laboratory Lead: Olga Simek
- Explainable Machine Learning for Decision Support
    - PI: Professor Aleksander Madry (EECS)
    - Co-PIs: Professor Arvind Satyanarayan (EECS), Professor Antonio Torralba (EECS)
    - Lincoln Laboratory Leads: Theodoros Tsiligkaridis, Steven Gomez, and Kevin Nam
- Robust AI Development Environment

  - PI: Professor Asu Ozdaglar (EECS)
  - Co-PIs: Professor Aleksander Madry (EECS), Professor Pablo Parrilo (EECS)
  - Lincoln Laboratory Lead: Olivia Brown
- Objective Performance Prediction & Optimization Using Physiological and Cognitive Metrics
  - PI: Professor Thomas Heldt (EECS)
  - Co-Is: Professor Vivienne Sze (EECS), Professor Tamara Broderick (EECS)
  - Lincoln Laboratory Lead: Hrishikesh Ra
- Robust Neural Differential Models for Navigation and Beyond
  - PI: Professor Alan Edelman (Mathematics)
  - Co-PI: Christopher Rackauckas (Mathematics)
  - Lincoln Laboratory Lead: Jonathan Taylor
- AI-Enhanced Spectral Awareness and Interference Rejection
  - PI: Professor Greg Wornell (EECS)
  - Co-PI: Professor Yury Polyanskiy (EECS)
  - Lincoln Laboratory Leads: Binoy Kurien and Jarilyn Hernandez Jimenez
- AI Education Research
  - PI: Professor Cynthia Breazeal (Office of the Provost)
  - Lincoln Laboratory Lead: Diane Staheli
- Application of Coevolutionary Algorithms for DoD Complex Enterprises
  - PI: Una-May O’Reilly (CSAIL)

Toyota-CSAIL Joint Research Center

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—one that is motivating the research of the Toyota-CSAIL Joint Research Center, which was kicked off in September 2015. The center is in collaboration with the Toyota Research Institute (TRI) led by Gill Pratt.

The objective of the Toyota-CSAIL Joint Research Center is 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. The CSAIL researchers are working on new tools for collecting and analyzing navigation data with the objective to learn from humans; perception and decision making systems for safe navigation; systems that can handle difficult driving situations: congestion, high speed driving, and inclement weather; predictive models that can anticipate the behavior of humans and vehicles; more intelligent user interfaces; human support robots, whereby robots can work safely and seamlessly with and around humans in performing tasks, establish effective communication with humans and recognize human intent.

The initial five-year project has concluded, and CSAIL is now participating in University 2.0, a long-term collaboration with other universities, which launched in April 2021 and has final funding projected for April 2025. The projects and PIs active in the Toyota-CSAIL Joint Research Center during the AY2021 include the following:

- Task Driven Development of Nimble, Reactive, Rugged Hands
  - PIs: Ted Adelson (Brain and Cognitive Sciences), Sangbae Kim (MechE), Alberto Rodriguez (MechE), Wojciech Matusik (EECS), Pulkit Agrawal (EECS)
- How can we create superintelligent human-computer groups?
  - PIs: Thomas Malone (Sloan School of Management), Daniela Rus (EECS), Abdullah Almaatouq (Sloan School of Management)
- Learning Interactive and Responsive Driving
  - PIs: Sertac Karaman (AeroAstro), Daniela Rus (EECS)
- Mini City Road Challenge
  - PIs: Daniela Rus (EECS), Evangelos Theodorou (Georgia Tech), James Rehg (Georgia Tech), Sertac Karaman (AeroAstro)
- Scalable Self-Supervised Learning for 3D Scene Understanding
  - PIs: Justin Solomon (EECS), Greg Shakhnarovich (Toyota Technological Institute at Chicago)
- Physical and Functional Inductive Biases for Visual Representation Learning
  - PIs: Joshua Tenenbaum (Brain and Cognitive Sciences), Jiajun Wu (Stanford), Fredo Durand (EECS)

Wistron-CSAIL Research Collaboration

Good health—both mental and physical—is one of the most pressing present-day 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 so 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 we consume 11–14 hours of technology each day. And this often involves multitasking, which in turn retrains our brains, reduces concentration, and increases stress (e.g., studies show that the brains of heavy technology users show similar patterns to those of substance addicts). Finding ways to reduce stress and technology's negative impact on a workforce is critical to our well-being in the future.

The multiyear research program between Wistron and CSAIL 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: How to 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 to develop systems that deliver appropriate lighting? How to develop systems that re-engineer email? How to develop algorithms that can help with information overload? How to use computing and communication in support of individual and community well-being? How to 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 this program focuses broadly on the computer and communication platforms. The second thrust focuses on using these novel platforms to promote healthier living. More specifically, the three projects that are currently active in the Wistron-CSAIL Research Collaboration are:

- Individualized Prediction and Interpretation of Risk: Predicting Trajectories of Chronic Disease and Recovery
    - PIs: Polina Golland (EECS) and Peter Szolovits (EECS)
- Using Machine Learning to Build Better Clinical Support Tools
    - PI: John Guttag (EECS)

CSAIL-Microsoft Trustworthy and Robust AI Collaboration

The Trustworthy and Robust AI Collaboration between CSAIL and Microsoft Research has been working towards fostering advances on robustness and trustworthy AI, which spans safety and reliability, intelligibility, and accountability. The collaboration seeks to address concerns about the trustworthiness of AI systems, including rising concerns with the safety, fairness, and transparency of technologies.

The collaboration leverages the mutual interests of Microsoft and MIT CSAIL on achieving AI systems in the autonomous, semi-autonomous, and collaborative realms but is centered on the vision of extending and augment the abilities and intellect of people. The research engagement includes funded projects between Microsoft researchers and MIT professors and students. The currently active projects are:

- Safe Online Reinforcement Learning in Networked Systems
    - PI: Mohammad Alizadeh (EECS)
- Machine Learning with Theoretical Guarantees
    - PI: Tamara Broderick (EECS)
- Exploration of Robust Machine Learning for High-Stakes Predictions
    - PI John Guttag (EECS)
- Explainability and Interpretability at Scale
    - PI: Stefanie Jegelka (EECS)
- Uncertainty and Robustness at Scale
    - PI: Aleksander Madry (EECS)
- Robustness meets algorithms
    - PI: Ankur Moitra (Mathematics)
- State-based Approaches for Verifying and Testing Neural Networks
    - PI: Martin Rinard (EECS)
- Compression for Interpretability at Scale
    - PI: Daniela Rus (EECS)
- Distributed, Private and Efficient Machine Learning
    - PI: Vinod Vaikuntanathan (EECS)
- Off-policy Evaluation for Risk-Aware Autonomous Systems
    - PI: Cathy Wu (Civil and Environmental Engineering)

In addition, in the reported academic year, this collaboration hosted two colloquium events:

- December 15, 2020, “What does robustness mean in ML?”
  Speakers: Ludwig Schmidt (University of California at Berkeley/University of Wisconsin) and Jacob Steinhardt (University of California at Berkeley)
  Moderators: Jerry Li (Microsoft Research) and Ankur Moitra

- May 24, 2021, "Causal inference and sequential decision-making"
  Speakers: Susan A. Murphy (Harvard) and Jonas Peters (University of Copenhagen)
  Moderators: Emre Kiciman (Microsoft Research) and Cathy Wu

As planned, this collaboration concluded at the end of AY2021.

Quanta-CSAIL Research Collaboration

This project is based on using machine learning to further our understanding of the following:

- What makes people more or less susceptible to different kinds of infections
- How different kinds of infections are spread
- How to predict which infected patients are most likely to experience severe symptoms
- How to design and test interventions that reduce the spread of infectious diseases

Our initial experiments have been in the area of healthcare-associated infections. These are infections that are spread within medical settings. In FY2021, our focus has shifted towards community acquired infections, including Covid-19. We made methodological progress on two critical issues: understanding and mitigating the impact of limited and biased testing, and combining imaging data with data extracted from electronic health records (EHR).

The best way to limit the spread of an infectious disease is to avoid transferring pathogens from those who are carrying the disease from those who are vulnerable to contracting the disease. A major challenge in doing this is identifying latent carriers (as we have seen with spread of Covid-19). One of the ways that machine learning algorithms can help is by building models that predict who is likely to become infected, making them good candidates for preemptive interventions. In this work we asked: can we build reliable infection prediction models when the observed data is collected under limited, and biased testing that prioritizes testing symptomatic individuals? Our analysis suggests that when the infection is highly transmissible, incomplete testing might be sufficient to achieve good out-of-sample prediction error. Guided by this insight, we developed an algorithm that predicts infections.

Large influxes of hospitalized patients during the Covid-19 pandemic caused significant and acute disruption of the healthcare system. In this work, we developed a method using machine learning to augment existing deterioration indices with chest radiographs. We applied the approach to predict deterioration of patients hospitalized with Covid-19, augmenting two existing EHR-based deterioration indices. While tested on the clinical problem of Covid-19, this approach has broader applicability to risk models that could be augmented with imaging data, especially for emergent diseases.

Industrial Outreach

CSAIL Alliances

The CSAIL Alliance Program

The CSAIL Alliance Program (CAP) is a gateway into the lab for industry, organizations, and governmental institutions seeking a closer connection to the work, researchers, and students of CSAIL. The program provides 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 industry and organizations across the globe.

Levels of Membership

The CAP program provides a proactive and comprehensive approach to connect members to the whole lab—all 60 research groups spanning robotics, natural language processing, networks, databases, cryptography, web science and more. CAP has three levels: student engagement, which focuses on connecting with students and post docs for career opportunities; affiliate, which provides lab visits, access to the annual meeting, recruiting assistance, research briefings and professional education discounts; and partner, which includes all of the benefits of the affiliate level as well as more expanded options with added access to research initiative meetings, custom faculty-led seminars and expanded recruiting options.

Member companies

Currently there are over 100 member companies, including CSAIL startups and global brands such as Apple, Cisco, Dell EMC, Facebook, Google, Lenovo, and Microsoft. Members are headquartered in North America, South America, Europe, and Asia and represent a wide variety of industry verticals.

Online Courses and Professional Development

CSAIL Alliances also produces and manages online professional development courses in partnership with MIT's Professional Education, MIT's Office of Digital Learning (ODL), edX, Get Smarter, and Harvard Extension School.

In FY2021, we began a new partnership with Emeritus and will be launching our first course on AI and Automation for the Enterprise in FY2022. We have also partnered with the US Airforce to bring professional development courses to their members.

The following is a list of the programs to date, a brief description, number of offerings to date and total enrollments to date. Total enrollment is now over 36,000 online learners:

CSAIL Online Courses and Professional Development Program Offerings, FY2021

Course title	Description	Courses offered	Enrollment
Artificial Intelligence: Implications for Business Strategy	Offered in partnership with the Sloan School of Management, this course focuses on the organizational and managerial implications of AI technologies	33	17,092
Machine Learning: Implementation in Business	This course is offered in partnership with the Sloan School of Management and aims to demystify machine learning for the business professional—offering a firm, foundational understanding of the advantages, limitations, and scope of machine learning from a management perspective.	13	1,328
HCI: Human Computer Interaction for User Experience Design	The six-week course was produced in partnership with Get Smarter and includes eight CSAIL researchers who review a host of cutting-edge HCI concepts including voice activated and spatially aware computers as well as speech and vision tools.	16	1,037
Tackling the Challenges of Big Data (not offered FY2019)	Survey state-of-the-art topics in Big Data, looking at data collection (smartphones, sensors, the Web), data storage and processing (scalable relational databases, Hadoop, Spark, etc.), extracting structured data from unstructured data, systems issues (exploiting multicore, security), analytics (machine learning, data compression, efficient algorithms), visualization, and a range of applications.	10	11,431
Tackling the Challenges of Big Data—Taiwan (not offered FY2018)	The original course translated into traditional Chinese.	1	1,296
Tackling the Challenges of Big Data—Ilumno	The original course translated into Spanish and Portuguese, offered through Ilumno in collaboration with universities in South America.	3	1,032
Internet of Things: Roadmap to a Connected World—Ilumno	The original course translated into Spanish and Portuguese offered through Ilumno in collaboration with universities in South America	3	608
Cybersecurity: Technology, Application & Policy (not offered FY2019)	Six-week online course provides a holistic look at cybersecurity technologies, techniques, and systems.	8	3,199
Introduction to the Challenges & Opportunities of Big Data, the Internet of Things and Cybersecurity	A semester long course combining our Big Data, IOT and cyber courses offered for credit through Harvard Extension School.	6	551
Startup Success: How to Start a Technology Company in Six (Not So Easy) Steps (not offered FY2019)	This course discusses the lessons learned by Michael Stonebraker and Andy Palmer during their startup endeavors over a 30-year period. The lessons are distilled into six steps that any entrepreneur can follow to get a company going. Topics include the generation and assessment of ideas, the challenges of building a prototype, the recruitment of a talented team, the closing of the first financing round, and pursuing growth with the right business leadership.	2	359
Internet of Things: Roadmap to a Connected World	The course introduces both the broad range of IoT technologies and the most recent developments in the space.	8	3955
Total		**103**	**36,493**

SystemsThatLearn@CSAIL

The goal of SystemsThatLearn@CSAIL is to accelerate the development of systems and applications that learn. We intend to accomplish this goal through combining our expertise in systems and ML to create new applications for understanding complex relationships unearthed by analyzing the avalanche of data available today. Presently, however, software systems that incorporate machine learning are hard 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 (and sometimes daily) maintenance to ensure that their predictions and behavior continue to be accurate and useful. Integrating ML systems into traditional enterprise architecture, testing, and deployment processes are too complex, partly due to organizational silos that exist between systems engineers and data scientists. In application, many problems in large-scale software systems involve optimizations that benefit from predictions; such as scheduling, compilation, query planning, routing, data cleaning, and congestion control. Today, it is hard to apply machine-learning tools to design this type of system software.

Our approach to designing, training, and deploying humanlike will focus on the following four areas of investigation:

Heterogeneous architectures

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

Predictable composition

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

Distributed execution

In terms of the underlying infrastructure, complex machine learning methods also demand considerable parallel resources to train effectively. Once trained, models may be deployed either on massive parallel infrastructures (e.g., data centers) or may have to be reduced and distributed back to the heterogeneous components to be utilized where 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 offer 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. This includes, among other things, safeguards and ways of containing learning functions. In addition to building better systems for machine learning, we believe 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.

SystemsThatLearn@CSAIL is a large, multi-PI research program 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 to help data scientists and engineers understand their models, scalably train them, monitor their results, retrain models efficiently
- Useful models focusing on efficiently deploying models in distributed and datacenter settings, reusing and redeploying models, as well as creating development environments good for training and deployment
- Developing heterogeneously deployable models, i.e., models that can be decomposed across heterogeneous devices, or lower fidelity models that can run on sensors or smartphones and also on more powerful servers as well as developing models that are more interpretable.
- Tools for statistical monitoring and performance prediction 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. On March 29, 2017, we launched this initiative with founding members: BT, Microsoft, NOKIA Bell Labs, Salesforce, Schlumberger, and ScotiaBank. Additional members added in FY2018 include BASF, ElementAI, EY, and JP Morgan Chase. In FY2019, Facebook was added as a member. In FY2020, SystemsThatLearn@CSAIL gave 12 total grants worth $1,155,000 in funding. SystemsThatLearn@CSAIL entered its fourth year and provided $561,000 in funding over 10 grant awards. Since 2017, this initiative has provided $2,591,000 in funding to 47 research projects.

FinTech@CSAIL

Financial Technology (FinTech) is disrupting many aspects of financial services, banking, and insurance, among other industries. Not only will infrastructure and operations be disrupted, but new technologies, business models, services and even industries will be launched. FinTech holds promise not only for verified transaction systems such as block-chain, but also for technologies involving AI, security, data analytics/value extraction, machine learning, trust verification, risk management and privacy advances as well.

The Financial Sector has many unique attributes and at the core of a company's success in this sector is trust, security, value, and efficiency. Current technology roadmaps aren't perceived as providing sufficient guidance for FinTech, 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, disruptive technology development and start-up companies that are reinventing global financial services. Through FinTech@CSAIL, 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, gain new data insights, and improve security.

The breadth of research at CSAIL uniquely positions the lab to address a wide variety of challenges in the space. FinTech@CSAIL will include 15 researchers of MIT CSAIL, who have pioneered the fields of secure computation, ML, AI, data analytics and risk management. The goal is to advance the state-of-the-art in collaboration with select industry partners to address the hardest problems facing the finance industry today.

Through the rigorous research of our faculty coupled with our tradition of collaborating with industry, FinTech@CSAIL will address relevant business problems with long-term vision. Additionally, FinTech@CSAIL will draw from across the lab as well as other focused research initiatives in Cybersecurity@CSAIL and SystemsThatLearn@CSAIL, all of which come together for a very 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 the following:

- New approaches to efficient shared public ledger systems and digital currency
- Technologies to provide secure, multiparty computation as well as secure and private data extraction
- Advanced data analytics to apply to risk management and prediction
- Scalable, trusted systems
- Security of ML and AI systems
- ML and AI in automatic advising, compliance, and augmented assistance
- Applying natural language processing in the context of financial contracts
- Speech recognition applications
- Security of legacy systems
- Efficient processing and automation of tasks
- Anonymization of data and privacy

- Movement of datasets and sharing data sets securely
- New lending and payment technologies
- Addressing the changing role of banks as new technologies form the landscape of the future of the financial industry

FinTech@CSAIL was launched on July 18, 2018. The initiative is led by Professors Andrew Lo, Silvio Micali, Gary Gensler, and Randall Davis. Founding members include: NASDAQ, Ant Financial, Citigroup, Ryan Software, The London Stock Exchange Group, and Ripple Labs. State Street Bank, the Canadian Imperial Bank of Commerce, the Bank of International Settlements, Fidelity, Consensys, Itau, Bank Negara Malaysia, and Bank of New York Mellon have also joined. In FY2021, FinTech@CSAIL provided $950,000 in discretionary funding for 24 CSAIL PIs to sponsor 13 proposals.

MachineLearningApplications@CSAIL

Launched in September 2020, the MachineLearningApplications@CSAIL Initiative focuses on applications of the latest ML technologies and research on the resolution of current challenges limiting the abilities of ML and professional development that will help prepare a company's workforce for this digital transformation.

Many companies are unsure of how, where, or if they should leverage ML. Awash in data, they are looking to turn that data into intelligence that drives increasingly efficient processes. The valuable insights and impact across all functions from sales, marketing and customer engagement to logistics, cost control, fraud detection, security and more can be transformational.

Organizations who know how to leverage and integrate ML across their business will have a competitive advantage. All industries including retail, food and beverage, travel and tourism, household goods, construction, fashion, agriculture, manufacturing and packaging, education, pharmaceutical, health care and more will all benefit from the latest ML technologies.

MachineLearningApplications@CSAIL works both the research ecosystem and online learning to address the following questions:

- How can machine learning be leveraged for additional insights but with outcome guarantees or provability?
- How can organizations analyze more complex data sets?
- How can the results be trusted?
- How can training models be updated with new data to keep the ML systems operating most efficiently?

This new initiative is led by CSAIL Director Professor Daniela Rus. Current members include: Ahold Delhaize, Arrow Electronics, Cisco, NEC Labs, CapitalOne and SAP. In FY2021, MachineLearningApplications@CSAIL provided $590,000 in discretionary funding for 19 proposals and supported 18 CSAIL PIs. This funding included eleven $10,000 awards to support UROPs within CSAIL.

The Future of Data, Trust, and Privacy

Launched in April 2021 in collaboration with MIT Internet Policy Research Initiative (IPRI), the Future of Data, Trust, and Privacy is the latest CSAIL strategic research initiative. This strategic research initiative brings together founding members: American Family Insurance, Capital One, MassMutual and Microsoft with state-of-the-art MIT computer science research and world-renowned experts to craft technologically-informed public policy and create new privacy-preserving tools to address today's data privacy challenges. Participants will seek to understand the implications of new laws such as the General Data Protection Regulation and to lead a global dialogue with policymakers, civil society, and industry leaders to shape the future of privacy and data governance. Technical research will focus on the following areas:

- Database Systems: Develop new data management architectures to provide enterprises with purpose management, provable delete and automated audit accountability tools for managing personal data according to existing and future privacy laws and governance models.
- Applied Cryptography: We will bring together the cryptographers and public policy experts to expand cryptographically powered data handling techniques (such as Secure Cyber Risk Aggregation and Measurement) to enable uses of personal data while limiting privacy risk
- AI and ML: Develop privacy preserving, trustworthy ML systems meeting global legal requirements and providing explanation and bias assessment.
- Data Portability and New Information Architectures: Design new protocols for managing personal data flow across APIs to enable support for data portability requirements while maintaining usage limits and accountability.
- Human-Computer Interaction: Apply rigorous HCI research methodologies to understand the impact of various privacy policy environments on user behavior and learn when the user experience is producing chilling effects. This research will inform both services design and policymaking.

The initiative will provide a strategically managed forum for dialogue amongst MIT researchers, policymakers, industry consortium members, and civil society partners.

The Future of Data, Trust and Privacy will be co-led by Daniel Weitzner and Professor Srini Devadas. Weitzner's research pioneered the design of accountable systems as a new approach to privacy and investigated the interaction between cryptographic technology and surveillance law, while Devadas's current research areas are computer architecture, computer security, and applied cryptography. These leaders in their respective fields will bring together a deep technological and policy understanding to the critical issues in data governance and management.

Internet Policy Research Initiative

Communication and information networks have become fundamental to our increasingly digital economy and society. Despite this importance, the technologists and policymakers who play key roles in supporting the transition to these networks approach issues from different perspectives and often do not speak the same technical languages. This disconnect can lead to uninformed policymaking and misdirected research efforts. As such, there is a pressing need to bridge the gap between technical and policy communities.

The mission of the Internet Policy Research Initiative (IPRI), an Institute-wide initiative, is to work with policymakers and technologists to remedy this issue and increase the trustworthiness and effectiveness of interconnected digital systems, like the Internet. We accomplish this via a three-pronged approach: targeted engineering and public policy research, educational programs for students and policymakers, and outreach programs to build policy communities that facilitate communication.

IPRI's Research Efforts Cover Six Categories

Cybersecurity

IPRI's cybersecurity research focuses on the technical and policy aspects of cybersecurity issues as they relate to the communication networks and software systems affecting the global society and economy. This multidisciplinary research area encompasses encryption policy, accountability, cryptography, data sharing, securing core economic and social infrastructure, measuring cyber risk, and more. Major projects include the following:

- The PACT (Private Automated Contact Tracing) project, a joint collaboration between CSAIL, IPRI, Massachusetts General Hospital Center for Global Health, MIT Lincoln Laboratory, and more. PACT aims to enhance contact tracing in pandemic response by designing exposure detection functions in personal digital communication devices that have maximal public health utility, while preserving privacy
- An interdisciplinary project called Secure Cyber Risk Aggregation and Measurement (SCRAM) which designs and builds cryptographic tools and platforms that allow us to measure cyber risk more accurately, without putting participant data at risk of discovery by other participants or the platform host
- The Future of Data, Trust, and Privacy, as detailed earlier in this report

AI Policy

Artificial intelligence and machine learning technologies are becoming increasingly prevalent in not only advertising and research, but also in traditionally regulated spaces such as healthcare, finance, transportation, and employment. The AI Policy team at IPRI is focused on increasing the trustworthiness of AI and ML systems by enhancing their explainability and accountability. Current research areas include studying the role of AI in financial decision making, increasing access to new training data sets with policy, working with stakeholders on AI principles, and shaping global Internet policymaking via policymaker engagement and informing the public debate. As part of this work, IPRI co-hosted the first AI Policy Forum summit with the MIT Schwarzman College of Computing on May 6–7, 2021. The summit was a collaborative virtual gathering of the AI Policy Forum task forces to discuss their progress toward equipping high-level decision-makers with a deeper understanding of the tools at their disposal—and trade-offs to be made—to produce better public policy around AI, and better AI systems with concern for public policy.

Privacy

Work also focuses on privacy policy and its critical role in trustworthiness. The Privacy group has published work on topics like privacy and security in home assistants, exposure minimization, and the data-sharing practices of smartphone apps. Current projects include the development databases that are privacy-aware.

Networks

The Advanced Network Architecture (ANA) group works to understand and shape the future of the internet. They achieve this goal with the understanding that the future of the Internet is defined by the economic, social, regulatory, legal, and political concerns involving the Internet. As such, the ANA group is organized around five themes: Internet architecture, Internet security, Internet economics, Internet policy, and network management. In 2018, the ANA group had its first major release of data about interconnection congestion on the Internet.

Decentralized Web

The Decentralized Information Group focuses on data and systems governance (primarily on the web) and explores both policy and technical issues. Current projects include a decentralized privacy-preserving platform for clinical research, evaluating the trustworthiness of autonomous systems, studying the relationship between privacy and machine learning, developing explanations for complex machines and models, securely aggregating distributed data, and developing smart contracts for data sharing.

App Inventor

The core goal of the MIT App Inventor team is to empower young people to develop useful apps that serve as novel digital solutions to the problems they face in their lives, communities, and world.

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. Nearly 450 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.

Over the last 18 months, the world has accelerated its movement from physical interaction to virtual. This is exciting, humbling, and troubling that this momentum resulted from a public health nightmare. Whatever the cause of the shift to virtual, it will only continue to accelerate. The web has become a key technical infrastructure for global society. The pattern of solving problems that are critical to society, getting them to scale to web level, ensuring interoperability, and solving challenging problems of security, privacy, performance, accessibility, and internationalization is the sine qua non for society and the raison d'etre of the new W3C going forward.

Work on WebRTC occurs in real-time communications on the web, and we are looking at changing the web advertising model so it is less intrusive than the currently used privacy invading methods. We are streamlining e-commerce, enabling immersive applications, looking at the exploitation of high speed, low latency networks (e.g. 5G), and streaming in media and entertainment. The web, which sits at the interface between what the network provides and what human imagination craves, has a limitless set of accelerating challenges that we must address in an agile fashion.

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 in unison about how HTML5 is the cornerstone for this platform. The full strength of the platform relies on many more technologies that the W3C and its partners are creating, including cascading style sheets. The W3C's real-time communications spec is becoming more central to society as physical meetings are replaced by virtual real-time meetings over the web.

In recent years, publicly noted security and privacy breaches have resulted in unprecedented attention to fixing web security and privacy. W3C addresses that both with specific solutions (such as the Web Authentication specification that authenticates users without passwords), and by conducting reviews of every W3C standard for security and privacy. There is new focus on how to support advertisers without the current approaches of having so much personal information available to advertisers. The growth of e-commerce has focused new attention on standardizing payment and e-commerce approaches. With immersive technologies, there is a strong focus on web solutions for virtual reality and augmented reality.

The growing impact of the Web is also driving W3C to expand its agenda and the size of its community. W3C launched Community and Business Groups in 2011. After a decade roughly 12,000 people participate. 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, advertising, personal communications, and financial services. This has caused a diversification in the membership of W3C, and also has enriched the technical agenda to address new technical issues that arise. For example, web browser companies, credit card companies, and other fintech stakeholders are using W3C to streamline web payments through a browser.

Research Highlights

In addition to the large-scale collaborative projects and center research, numerous individual and multi-investigator projects are under way.

CSAIL researchers had major research discoveries across all areas of computing. Our work on advancing computing, making it more powerful and more pervasive, opens many opportunities to propel science, create new businesses, protect the planet, understand life, improve our cities and enhance our well-being and quality of life. CSAIL colleagues made important advances in cybersecurity, demonstrating the vulnerabilities of online voting systems and creating secure machine learning systems. In computer architectures, cross-campus teams explored the post-Moore's Law era of computing. Machine learning advances allowed our scientists to develop new antibiotics that could help treat antibacterial infections. Others created systems for home settings that analyze in-home appliance use to better understand people's day-to-day health habits. Researchers have even explored the realm of imbuing algorithms with human-like traits of curiosity. In database systems, PIs are developing systems that use machine learning to automatically reorganize a dataset's storage layout based on the types of queries that its users make. There also continue to be groundbreaking findings in theory and systems, from proving lower-bounds on SAT algorithms to creating new algorithms for incremental single-source shortest path problems.

A more detailed sampling of the work is highlighted below:

Large-scale Reconfigurable Optical Networks

Led by Associate Professor Manya Ghobadi, large-scale optical networks are the foundation of modern online services. As the world is recovering from the Covid-19 pandemic, there is a vital reliance on online service providers to deliver high bandwidth, low latency, and high availability for emerging workloads, such as remote video calls, augmented reality, machine learning, and health care. However, the design of today's network infrastructures, both in datacenters and in wide-area networks, still follows the telephony model where network operators (such as AT&T, Comcast, Google, Facebook, and Microsoft) treat the physical layer of networks as a static black box with no reconfigurability. As a result, the network infrastructure is provisioned to carry the worst-case traffic demand under all plausible failure scenarios, making it excessively inefficient and prohibitively expensive.

To solve this problem, we have developed several new paradigms for large-scale dynamic reconfigurable networks. Our work is based on an unconventional approach to dynamically change the optical topology of networks to adapt to applications' traffic demands and recover from failures. Reconfiguring the optical layer of a network is a major departure from the common practice that has prevailed for decades. To address this challenge, we have developed a series of solutions that significantly improve the cost and throughput of both datacenter and wide-area networks by reconfiguring the optical connections dynamically, in a manner similar to wireless networks. Our work covers the spectrum, from recovering several tera-bits-per-second of capacity during lengthy fiber cut events to accelerating ML workloads with reconfigurable optical topologies.

Tensor Holography

Led by Professor Wojciech Matusik, bringing photorealistic true 3D viewing experience to virtual and augmented reality (VR/AR) has been an ultimate goal for researchers in both industry and academia. Existing VR/AR headsets have yet to topple TV or Smartphones as the go-to devices for video viewing because the 3D experience was faked by feeding the eyes with a pair of parallax images. This causes nausea and eye strain after long-time use since the eyes are still staring at fixed-distance 2D displays.

Matusik's group aims to change this state of affairs. The solution for better 3D visualization lies in a 60-year-old technology remade for the digital world: holograms. A hologram records the light waves emitted by the entire 3D scene and can be replayed in sequence like a standard video. Yet, the tremendous computation cost to simulate and optimize a high-quality digital hologram was deemed intractable for current-generation computational devices. The group's latest work published at *Nature* has changed this common belief with the first deep-neural-network-driven hologram simulator. It astoundingly improved the computational efficiency by more than 100 times with even better visual quality, the first time allowing real-time computational on a consumer-level personal computer and even interactively on an iPhone. This breakthrough can potentially change the landscape of AR/VR for the next ten years by drawing significantly more resources into holographic research and pushing it into products.

Beyond volumetric displays, this new technology has far-reaching impacts in other fields such as:

- Additive manufacture: will allow an instantaneous single-shot complex 3D shape forming by projecting the entire 3D volume at once.
- Microscopic manipulation: will 3D optical and acoustic tweezers with a significantly increased number of foci and more fine-grained control.
- Holographic microscopy: will allow real-time simulation of physically recorded digital holograms from biological samples.

The promises shown by this technology have drawn attention from industry partners such as Sony and NCSoft, who provide additional fundings to support our further pushes. The group is currently actively improving this technology with other key features such as an embedded hologram lens for vision correction to get rid of vision correction eyeglasses and pupil steering to support more significant eye movement. Our goal is to perfect this technology to clear all possible barriers for bringing it into everyone's daily life.

Health: Algorithm Reduces Unnecessary Use of Antibiotics for Urinary Tract Infections

Led by Associate Professor David Sontag, urinary tract infections (UTIs) affect half of all women, who typically must receive an antibiotic to resolve the infection. Given the significant growth in antibiotic resistance, clinicians face a difficult decision of which antibiotics to prescribe. Clinicians must choose an antibiotic that will resolve the infection, at the same time as being good antibiotic stewards, which necessitates minimizing the use of broad-spectrum antibiotics.

In collaboration with Mass General Brigham, we developed a ML-based decision support tool to suggest to clinicians which antibiotic to prescribe, based on both previous medical records from the patient and population-level data such as recently observed levels of resistance. Our algorithm was trained on data derived from electronic medical records of all women with uncomplicated UTIs from 2007 through 2013. When applied to a test cohort of 3,629 patients presenting between 2014 to 2016, we found that it suggested inappropriate antibiotics 18% less often than clinicians, at the same time as achieving a 67% reduction in the use of broad-spectrum antibiotics.

Laboratory Sponsored Activities

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 8,000 comments and questions, as well as more than 300,000 page views.

CSAIL Media Outreach

CSAIL has a combined online following of 340,000 users across Twitter, Instagram, Facebook, LinkedIn, and YouTube; with growth in media coverage and viewership at a 900% increase in media hits since 2013. The amount of CSAIL video that YouTube users have watched is now more than 166,900, which is equal to more than 19 years. Our YouTube viewership comparison with MIT News is as follows: CSAIL—17 videos, with an average of 23,400 views; MIT News—55 videos, with an average of 21,966 views (median: 10,466). Consistent media coverage in various top tier outlets includes BBC, CNET, CNN, FastCompany, IEEE Spectrum, TechCrunch, and Yahoo.

CSAIL Hosted Lecture Series

Dertouzos Distinguished Lectures have been a tradition since 1976, featuring some of the most influential thinkers in computer science. Two speakers presented lectures during the AY2021 Dertouzos Distinguished Lecture Series:

- Senator Ron Wyden, Oregon. Fireside discussion of data privacy, September 30, 2020
- Professor Ross Anderson, University of Cambridge, "Infrastructure: The Good, the Bad and the Ugly," March 10, 2021

Hot Topics in Computing is a speaker series initiated by CSAIL in 2017, which convenes experts in computing to discuss emergent potential, perception, and problems associated with the proliferation of computation and machines. We held a much higher number of events (nine events in AY2021), to address coronavirus and Covid-19 pandemic-related topics:

- Professor James Collins, "Harnessing Synthetic Biology and Deep Learning to Fight Pathogens," September 23, 2020
- Timnit Gebru, Google Ethical AI, "Computer vision-who is harmed and who benefits?" October 8, 2020
- Kenan Sahin Distinguished Professor Charles Stewart, MIT, "Will the 2020 Election Be Safe and Secure?" October 20, 2020

- Fireside Talk with Eric Schmidt, November 17, 2020
- Hany Farid, University of California at Berkeley, “Creating, Weaponizing, and Detecting Deep Fakes,” December 2, 2020
- Fireside Chat with Neil deGrasse Tyson, December 9, 2020
- Stacey Gabriel, senior director of the Genomics Platform at the Broad Institute, “Viral diagnostics for SARS-Cov-2: Setting up testing at massive scale,” February 24, 2021
- Fireside Chat with Neil deGrasse Tyson, April 7, 2021
- David Carroll, associate professor of media design, The New School, “Data Quest-Repatriating My Data from Cambridge Analytica,” May 5, 2021

Organizational Changes

Professor Daniela Rus has continued in her roles as director of CSAIL and deputy dean of Research in the Schwarzman College of Computing. CSAIL director’s duties include developing and implementing strategies designed to keep CSAIL growing and evolving, fund raising, determining laboratory policies, and examining promotion cases.

CSAIL’s leadership team includes an associate director and 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 Armando Solar-Lezama were the FY2021 associate directors. Professor Solar-Lezama also served as COO, providing leadership and strategy for how we conduct our operations and initiatives, enabling the director to allocate more time to strategic planning.

The CSAIL executive cabinet met weekly to review and advise the director on policy, processes, activities within the laboratory, preparation for transitions related to the Scharzmann College of Computing. Members of the AY2021 executive cabinet included John and Dorothy Wilson Professor Edward Adelson, Professor Randall Davis, Senior Research Scientist James Glass, Professor Daniel Jackson, College of Computing Distinguished Professor of Computing Samuel Madden, Wojciech Matusik, Edwin Sibley Webster Professor Ronitt Rubinfeld, Andrew (1956) and Erna Viterbi Professor of Electrical Engineering and Computer Science and Director of CSAIL Daniela Rus, Distinguished College of Computing Professor Armando Solar-Lezama, and Professor of Biological Engineering and Computer Science Bruce Tidor.

The CSAIL enterprise services team manages lab operations. There are seven units: Administrative Assistants, CSAIL Alliance Program, Communications, Finance, Human Resources, Special Projects, and the Infrastructure Group. These report to the CSAIL COO on all operational matters. Carmen Finn is the assistant director for administration overseeing the finance and human resource operations; she also oversees the administrative assistants core and serves on the space committee. Lori Glover is managing director of Global Strategic Alliances, overseeing CSAIL Alliance Program and course collaborations. Peter Jones is assistant director of planning and initiatives, overseeing sponsored program management, strategic initiatives, and the Director Office staff. Jack Costanza is assistant director for infrastructure, overseeing information technology infrastructure and user support, and building operations; he serves on the space committee.

Bruce Tidor oversees the space committee and manages the allocation of space within CSAIL. The space committee also implements improvements to the facilities that will increase the quality of the environment for the laboratory's faculty, staff, and students. He also manages Covid Pass requests and approvals within CSAIL, and manages requests for one-time access in line with Covid policies.

New faculty starting during AY2021 include the following:

- Assistant Professor Henry Corrigan-Gibbs, July 1, 2020
- Assistant Professor Anand V Natarajam, September 1, 2020
- Assistant Professor Ashia Wilson, January 1, 2021

Faculty taking leave during AY2021 included:

- Silvio Micali, professional leave
- Polina Golland, sabbatical leave
- Vinod Vaikuntanathan, sabbatical leave
- Hari Balakrishnan, personal leave
- Shafi Goldwasser, personal leave

Awards and Honors

Our faculty and staff have achieved many awards and honors, including the following:

- Jacob Andreas: 2021 Faculty Innovation Award from Sony
- Hari Balakrishnan: 2020 Infosys Prize in Engineering and Computer Science; 2021 Distinguished Alumnus Award, University of California at Berkeley; and 2021 Institute of Electrical and Electronics Engineers Koji Kobayashi Computers and Communications Award
- Regina Barzilay: 2021 Association for the Advancement of Artificial Intelligence Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity; 2021 United Nations Educational, Scientific and Cultural Organization (UNESCO)/ Netexplo Award; and 2021 Wallace H. Coulter Lectureship Award by AACC
- Tamara Broderick: 2020 Ruth and Joel Spira Award for Distinguished Teaching; 2021 Office of Naval Research Early Career Grant; and 2021 Committee of Presidents of Statistical Societies Leadership Academy Award
- Henry Corrigan-Gibbs: 2020 ACM Conference on Computer and Communications Security "Test of Time" Award; 2020 Doctoral Dissertation Honorable Mention Award, ACM; 2020 Facebook Research Award

- Erik Demaine: 2020 Award for Excellence in Teaching, MIT
- Alan Edelman: 2020 Fellow, ACM
- William Freeman: 2021 National Academy of Engineering Member
- Manya Ghobadi: 2021 Best Paper Award at the Machine Learning and Systems Conference
- Shafi Goldwasser: 2021 L'Oréal Foundation and the UNESCO Laureate
- Polina Golland: 2021 American Institute for Medical and Biological Engineering fellow
- Piotr Indyk: 2020 co-director, Foundations of Data Science Institute, NSF Data Science Institute
- Stefanie Jegelka: 2020 Two Sigma Faculty Research Award; and 2021 Google Faculty Research Award
- David Karger: 2020 Lasting Impact Award, UIST
- Tomas Lozano-Perez: 2021 IEEE Robotics and Automation Award
- Nancy Lynch: 2020 distributed computing contributions award, IEEE; 2020 Sorbonne honorary doctorate; and 2020 Test-of-Time Award, CONCUR
- Sam Madden: 2020 Fellow, ACM
- Rob Miller: 2020 Lasting Impact Award, UIST
- Jonathan Ragan-Kelley: 2020 Intel's Outstanding Researcher Award
- Daniela Rus: 2020 Fellow, American Association for the Advancement of Science; International Joint Conference on Artificial Intelligence John McCarthy Award for AI; and MIT Excellence Award for Serving our Community
- Arvind Satyanarayan: 2021 ACM Conference on Human Factors in Computing Systems Honorable Mention; and 2021 Google Research Scholar Award
- Julian Shun: 2020 Ruth and Joel Spira Award for Excellence in Teaching, MIT School of Engineering; 2021 ACM Special Interest Group on Management of Data Research Highlight Award; 2021 Best Paper Award at the Proceedings of the International Symposium on Code Generation and Optimization; and Research Scholar Award
- Michael Stonebraker: 2020 C&C Prize for contributions to Relational Database, NEC C&C Foundation
- Ryan Williams: 2021 Frank Quick Faculty Research Innovation Fellowship, MIT Department of Electrical Engineering and Computer Science; and 2021 Society for Industrial and Applied Mathematics Review SIGEST Award

Key Statistics for AY2021

Personnel type	Count	Women %
Faculty	146	20%
Postdoctoral associate or fellow	98	20%
Principal research scientist	15	50%
Research staff	87	12%
Senior research scientist	7	40%
Administration, technical, and support staff	183	54%
Graduate students	500	28%
Undergraduate Research Opportunity Program students	484	48%
Visitors	3	0%
Total	**1,368**	**33%**

Daniela Rus
Director