CSAIL Alliances Annual Meeting

May 24-26, 2023

Key Takeaways

The Benefits of CSAIL Alliances

- Access to world-class MIT researchers
- Connections across subjects, departments, and research areas thanks to CSAIL's multidisciplinary approach
- Dedicated Client Relations Coordinator to tailor and facilitate a personalized course of action
- Discounts on MIT professional development courses
- Member-only conferences, talks, and symposiums (like this one!)
- Monthly written, video, and audio content on the CAP website
- Student engagement and recruiting opportunities
- Unique opportunities to connect with startups in the MIT ecosystem
- And more…

Paths of Engagement

What Generative AI Means for Business

Lori Glover, Managing Director, MIT CSAIL Global Strategic Alliances

- The majority of Americans believe AI will affect jobholders in the next 20 years, and an increasing number of businesses are beginning to adopt AI solutions.
- Generative AI can be used in numerous areas, such as product design, content generation, data augmentation, personalized recommendations and assistants, simulation and modeling, employee productivity, and anomaly or fraud detection.
- In order for society to fully embrace generative AI, better explainability measures need to be developed for increased understanding and trust.
- Despite the excitement, there are important concerns around generative AI to be aware of, such as the perpetuation of harmful biases, hallucinations, copyright concerns, and the risk of data breaches when putting proprietary information into a public model like ChatGPT.

A few of the ways our members are using Generative AI :

- ***Predictive modeling & estimating future demand.***
- ***Generating ideas, models, and concepts.***
- ***Increasing worker productivity, particularly in marketing, coding, and design.***

Startup Showcase

Steve Kommrusch

Researcher, Leela AI

- One of the challenges of creating a general AI is developing a common-sense, physical understanding of the world.
- To address this, Leela AI is adding transformer concepts to create AI that can construct knowledge in an active process.
- Currently deployed in manufacturing, Leela AI hopes to create an AI with common-sense intelligence based on how humans learn.

Gant Redmon

CEO, Hopara, Inc.

- Digital twins, such as those offered by Hopara, can be used to virtually represent real-world processes such as manufacturing plants, equipment, etc.
- Hopara is also creating Agile Digital Twins (ADTs) for time-sensitive applications.
- This research is based on the work done by Professor Michael Stonebraker, who wanted to design a way to navigate data the way we navigate with map applications.

Emanuel Zgraggen

Co-founder, Einblick

- Data processing often leads to issues around department siloing, collaboration bottlenecks, and communication challenges between experts and non-experts. To solve these problems, Einblick offers a multi-modal interactive and collaborative canvas that allows multiple users of various levels of expertise to work on the same data.
- Einblick's interface incorporates code, no-code, and natural language options to make data processing easier.

Designing the Next Generation Cloud Systems

Christina Delimitrou, Assistant Professor, MIT EECS

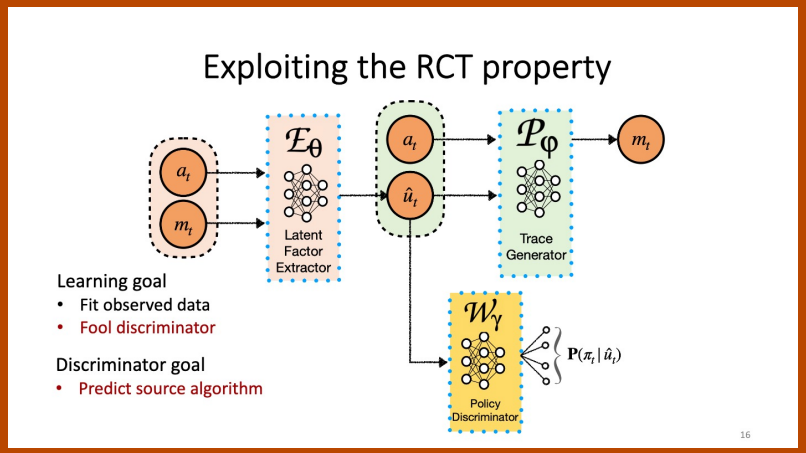
- Nearly all computing today involves the cloud. With current technological demands growing and Moore's Law slowing down, more specialized solutions are needed to handle the increasing complexity.
- Some areas Professor Delimitrou's group is exploring to design next-generation cloud systems are:
  - Applications and compilers;
  - Programming frameworks;
  - Cluster management & debugging;
  - Hardware design & acceleration.
- Fundamentally, she aims to develop ML-driven design and management for the cloud.
- Some of the solutions that have come from this research are:
  - **Seer**: a supervised and proactive debugging method that uses ML to identify an upcoming quality-of-service violation.
  - **Sage**: an unsupervised learning method that can explore the root cause of a quality-of-service violation.

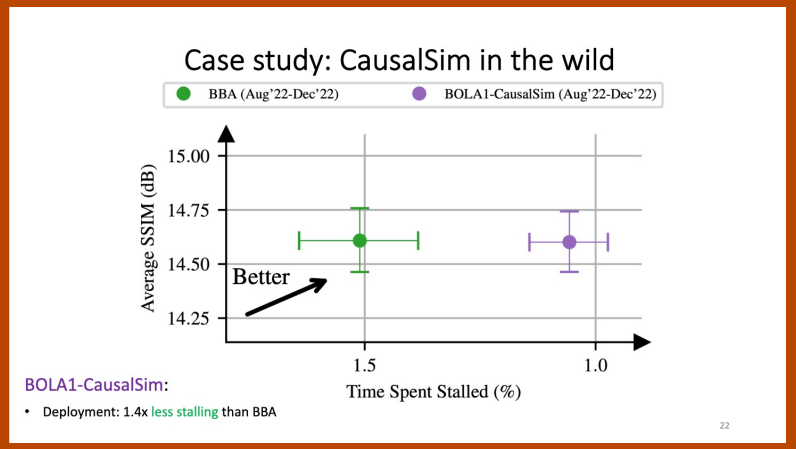
AI to Stimulate Computer Systems

Mohammad Alizadeh

Associate Professor, MIT EECS

- While useful to capture real system behavior and less complex than full system simulation, **trace-driven simulation** suffers the problem of biased outcomes due to the choices algorithms make during trace collection.
- To address this problem, Professor Alizadeh's group developed CausalSim, a framework for unbiased trace-driven simulation. CausalSim works by:
  - Using an initial randomized control trial (RCT) under a fixed set of algorithms;
  - Mapping unbiased trace-driven simulation to a tensor completion problem with sparse observations.
- Tested extensively on real-world data, CausalSim improves simulation accuracy, significantly reduces errors, and provides new insight into ABR algorithms.





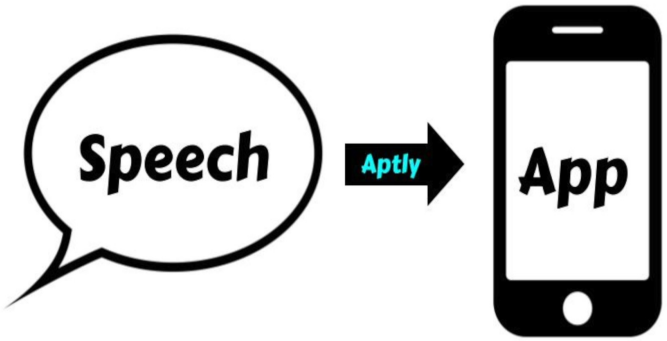
Exploring the Potential of Generative AI in K-12 Education

David Kim

Software Engineer, MIT App Inventor

- Started by Professor Hal Abelson, MIT App Inventor helps people around the world, including children and those without coding experience, design apps. App Inventor's 15 million users, ranging from 8 to 80 years old, have created 68 million apps in 195 countries.
- App Inventor is now working to add generative AI options in several ways, allowing users to:
  - Integrate chatbots into their apps;
  - Add image generation to apps;
  - Using Aptly, dictate what kind of app a user wants to create.
- App Inventor imagines a future in which AI acts as a translator between humans and computers to democratize programming and empower everyone to apply the power of generative AI.
- There's no way to stop children from using generative AI, so educators need to think about ways to insert AI into education, empowering students and preserving their abilities to interface with computer technology.

"Kids are people too!" Professor Hal Abelson



Welcome & CSAIL Updates

"AI will improve our lives in so many ways, some of which we have only begun to imagine."

- While there is much excitement around AI right now, it's important to consider the consequences of the technology we deploy in the world.
- CSAIL is careful to weigh the impact of what is being developed in our research labs.
- To celebrate everything CSAIL has accomplished, this summer on **June 27th** there will be an anniversary symposium and party commemorating 50 years since Project MAC was founded and 20 years since the AI Lab and the Laboratory for Computer Science merged to become CSAIL.

Daniela Rus

Director, MIT CSAIL

Software Development in the AI Age

Daniel Jackson, Associate Director, MIT CSAIL & Professor, EECS

- LLMs are already being used in the coding process, but they are imperfect assistants considering their fundamental limitations, namely that they don't tend to use common code and their solutions lack good structure.
- However, by exploiting familiar, reusable ideas and breaking larger projects into granular tasks, coders might be able to more effectively communicate with AI and generate better code.
- Professor Jackson presented the idea of constructing a repository of concept design principles—like a Wikipedia for developers—that would combine the power of generative coding models and human supervisors. This way, the AI agent could be pulling from a library of pre-existing code to build, for example, an app using established, familiar, and explainable segments.
- To help developers get started, Professor Jackson introduced his book, *The Essence of Software,* which aims to make the concept of these design principles accessible and actionable.

Future of Data, Trust, and Privacy

- In a time of dwindling consumer trust, transparency is critical.
- The market is shifting from expectations around procedural practices to technical standards.
- Some methods being explored right now are data traceability, accountability, and new frameworks such as open banking.
- The biggest challenge in policy creation is a lack of technical understanding in the policymakers and public, leading to concerns about fear-driven approaches that will stifle innovation.
- There are still many unsolved challenges in this field, which CSAIL researchers are hard at work addressing.

Bob Hedges, Chief Data Officer, Visa

Awah Teh, VP, Data Governance & Privacy Engineering, Capital One

John Mariano, Head of Data Aggregation and Data Science Technology, Fidelity

Professor Daniel Weitzner, Faculty Co-Director, MIT Future of Data, Trust, and Privacy

Professor Srini Devadas, Faculty Co-Director, MIT Future of Data, Trust, and Privacy

AI For Genomic Medicine: Disease Circuitry, Patient Subtyping, Drug Design

"This is as precision medicine as it gets."

Manolis Kellis, Professor, MIT Department of Biology

- Many of today's most challenging afflictions have thousands of loci influencing disease progression. However, if we can understand the disease circuitry, then it is possible to access the "knobs" of the disease and potentially intervene.
- Professor Kellis discussed discoveries coming out of his lab, including:
  - A single letter gene change that can lower obesity risk;
  - The ApoE4 gene target for Alzheimer's;
  - Research on identifying the genetic signatures that prevent patients from responding to immunotherapy.
- Some ways AI is proving useful in this domain are:
  - Training LLMs to understand protein domains in natural language (Example: AlphaFold);
  - Creating multi-tower models which encode protein structure & function in the same space and can translate between different domains;
  - Facilitating idea navigation between groups to prevent the siloing of information and encourage collaboration and connection.
- The more information we can tailor to and feed into these AI models, the closer we get to individualized precision medicine and a disease-free future.

Visual Computing

- Professor Wojciech Matusik began with the question: can computers beat humans at design? He went on to describe how robots can be represented graphically, how the design space can be conceptualized and mapped, and how simulation and reality can be bridged with neural nets and high data efficiency.

"We are not there yet, but we are extremely, extremely close." Professor Matusik

- Professor Polina Golland explored using machine learning to read X-Rays, both to aid in clinical diagnostics and help researchers access the vast amount of clinical data in old images.
- Associate Professor Justin Solomon introduced a lesser-known area of visual computing: **geometric data processing**. This includes shape identification, analysis, and representation with ML-style algorithms.

Wojciech Matusik, Professor MIT EECS

Polina Golland, Professor MIT EECS

Justin Solomon, Associate Professor MIT EECS

Robotics and Embodied Intelligence

"Right now, we have a real problem in the supply chain. All aspects of the supply chain can benefit from robots."
- Daniela Rus, Director MIT CSAIL

"Dexterous manipulation is making a lot of progress… we're seeing more and more capable robots."
- Russ Tedrake, Professor MIT EECS, Aero/Astro, MechE

- Professor Daniela Rus introduced the new Liquid Neural Networks being developed in her lab, which increase resiliency, stability, and expressivity.
- Discussing the challenges of creating general intelligence for robots, Professor Leslie Kaelbling explored how the metacognitive strategies humans use to perceive the world could be applied to robots, scaling algorithms down to focus on what's task-relevant, thus providing a temporary answer to the question: "how do you solve a big problem with a small brain?"
- Professor Russ Tedrake described how the paradigm of robotics has changed with better visual systems, behavior cloning, and simulation training. One remaining "big challenge" is getting robots to adjust to unforeseen situations.

Sustainable Computing

Daniela Rus

Director, MIT CSAIL

&

Jeremy Gregory

Executive Director, MIT Climate and Sustainability Consortium

- Climate change is already affecting our world—the last few summers have been the hottest on record—and energy-intensive computational solutions are only going to become more critical to our economies, workforce, and lives.
  - Data center consumption is estimated to increase 8-21% by 2030.
  - Training even a medium language model uses more than 600,000 pounds of carbon dioxide.
- However, technology itself might provide the solution.
  - AI can be used to optimize electricity usage, make transportation more efficient, monitor deforestation, preserve biodiversity, distribute food more effectively, minimize waste, etc.
  - New inventions could help capture carbon, block UV rays from reaching the earth, create liquid fuel from sunlight, and more.
  - Increasingly intelligent simulations of the climate, biodiversity, and ecology can help us understand the potential impact of various solutions and choose the best ones.
- Working together, academia, government, and industry can solve these problems through programs such as:
  - MIT's Climate & Sustainability Consortium;
  - CSAIL's upcoming sustainability research initiative.

MCSC | MIT Climate & Sustainability Consortium

Generative AI

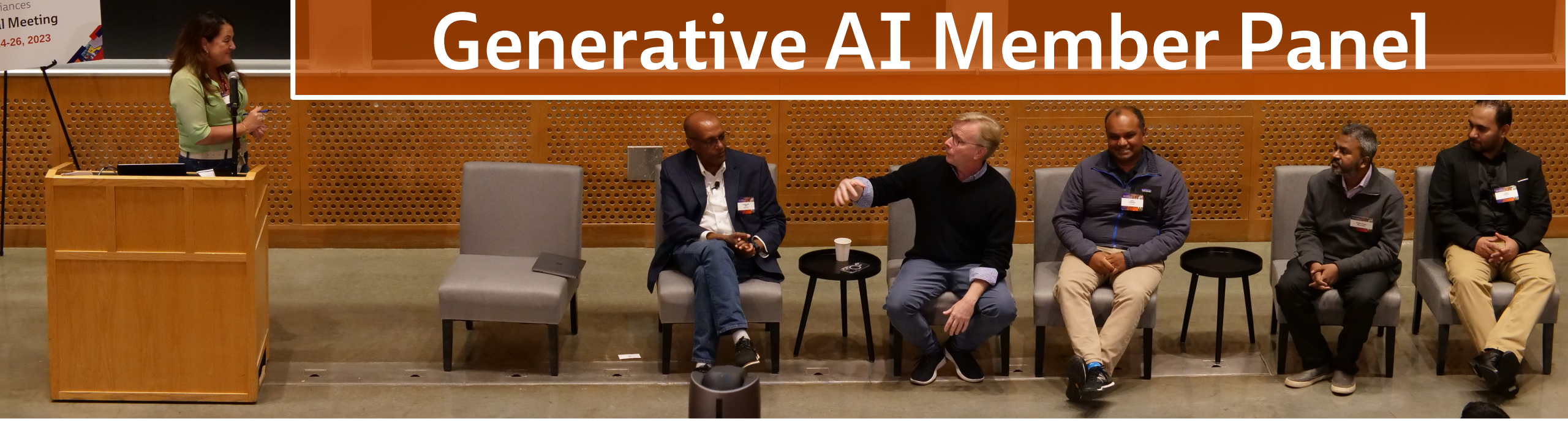
- Associate Professor Phillip Isola began the session discussing how synthetic data can applied toward generative AI models to speed up training, increase accuracy, and alleviate concerns around copyright and privacy. Some of the methods he discussed for using synthetic data were:
  - Generative Adversarial Networks (GANs)
  - Neural Radiance Fields (NeRFs)
  - Random Noise
- Addressing the safety of generative AI programs, Assistant Professor Dylan Hadfield-Menell spoke about brittle incentives and the dangers of designing AI systems without deeply thinking about the potential outcomes of a given system's objectives. He said, "**we should be wary about our ability to evaluate these systems**" and should instead be thinking about ways to optimize AI processes to actively identify unspecified qualitative harms rather than code them for straightforward obedience.
- Ending with a discussion of how LLMs can be retrained for other applications, Assistant Professor Yoon Kim discussed how generative AI models can be adapted and reused through more efficient training and deployment.

Dylan Hadfield-Menell, Assistant Professor MIT EECS

Yoon Kim, Assistant Professor, MIT EECS

Phillip John Isola, Associate Professor, MIT EECS

Generative AI Member Panel



(Panel left to right) Dagnachew Birru, Head of R&D, Quantiphi; Fred Goff, Co-founder and CEO, Jobcase; Sanji Fernando SVP AI & Analytics Platforms, Optum; Sadid Hasan, AI Lead, Microsoft; Adnan Masood, Chief AI Officer, UST

- Generative AI represents a foundational shift in how work will be approached, defined, and conceptualized.
- It's important for companies to think about how to bring this technology to the entire workforce, empowering employees to use these solutions productively.
- Generative AI models offer the ability to reduce cognitive loads and allow workers to do more with less time. This will be especially meaningful in fields with personnel shortages.
- There are ongoing concerns about the technology, especially the propensity of generative models to hallucinate, offer inaccurate information, or lack explainability.
- Generative AI models are expensive to implement with currently unknown ROI.
- The unclear future of AI regulation is something many companies are worried about, but holding to a high standard of privacy and consulting with academic centers like CSAIL can help businesses be prepared for the coming change.

Cybersecurity

- Principal Research Scientist Una-May O'Reilly explored how LLMs will impact cybersecurity by analyzing behavior to match for known patterns of attack, looking for anomalies, automating the understanding of cybersecurity knowledge to be accessible more quickly, helping with strategic analysis, training future cybersecurity professionals, and more. However, she also touched upon the duality problem, saying, "**in the wrong hands, language models can be used against me**."
- Assistant Professor Henry Corrigan-Gibbs presented some specific tools for protecting privacy, including **SimplePIR** and **DoublePIR** which facilitate private information retrieval in search engines and **Prio** for private data collection and aggregate statistics.
- Assistant Professor Mengjia Yan spoke about hardware vulnerabilities and the importance of considering the intersection of hardware and software when it comes to cybersecurity. She discussed the **PACMAN Attack** on Apple M1 processors that her group identified in 2022 and **Pensieve**, a security evaluation framework that automatically checks for microarchitecture vulnerabilities.
- Fundamentally, the task of keeping computer systems safe is an arms race which boils down to our collective ability to stay ahead of malicious actors. For that, Dr. O'Reilly explained that it's important to shift from reacting to the latest attacks toward thinking strategically with proactive approaches like cyber-hunting, supportive policy, education pipelines, etc.

"We need guardrails, but we don't want to shut things down so much that we don't get the benefits of this technology."
Professor Corrigan-Gibbs

Una-May O'Reilly, Principal Research Scientist, MIT CSAIL

Henry Corrigan-Gibbs, Assistant Professor, MIT EECS

Mengjia Yan, Assistant Professor, MIT EECS

Algorithmic Robust Statistics

Sam Hopkins, Assistant Professor, MIT EECS

- As our datasets about the world grows ever larger and more intricate, mathematicians turn to high dimensional statistics to handle this complexity.
- The issue is that high dimensional datasets are extremely noisy and messy, so to compute with them, one must assume there is a certain level of contamination from outliers, malicious actors, unclean data, etc. This leads to another problem, in that it is impossibly inefficient to check every datapoint, and challenging to even understand what an outlier might be in a given dataset.
- Algorithmic robust statistics aims to address this by designing models that offer reasonable running times for polynomial dimensions and allow for a certain level of “bad” data.

“Everything we try to do these days in modern statistics is extremely data-hungry and extremely compute-hungry”

The Future of Cryptography Panel

- Professor Vinod Vaikuntanathan introduced three major challenges of modern cryptography and some potential solutions to them:
  - The rise of quantum computing threatens current RSA cryptosystems, so cryptographers need to start exploring new encryption methods—such as lattice-based cryptography—now to be prepared.
  - The growing ubiquity of cloud computing opens users to new risks and vulnerabilities, which means we need to shift from secure communication to secure computation so that even if a cloud system is compromised, the information is not accessible.
  - There is much value in companies collaborating on issues like cybersecurity, but they might not want to share proprietary data. Innovative methods like homomorphic encryption and privacy-preserving data collection & search methods facilitates a future of secure collaboration and increased communal gains.
- Adjunct Associate Professor Yael Kalai went on to discuss the idea of verifiability on untrusted platforms and how a given output might be certified for correctness or completeness. Work at MIT has led to the development of SNARGs (Succinct Non-interactive ARGguments), which have since been deployed among companies.
- The “grand challenges” of verifiability going forward are providing SNARGs for ML models and shrinking SNARGs down so they take up less space but come with the same formal guarantees.
- Both professors agreed that the ideal end stage of cryptography is “making these technologies invisible and transparent to the end user.”

Yael Kalai

Adjunct Associate Professor, MIT EECS

Vinod Vaikuntanathan

Professor, MIT EECS

Connect with Us

CSAIL Alliances Website: cap.csail.mit.edu/

LinkedIn: www.linkedin.com/company/mit-csail

Twitter: @csail_alliances

YouTube: www.youtube.com/c/MITCSAILAlliances-1

SoundCloud: soundcloud.com/csail-alliances