C3.ai Digital Transformation Institute

ANNUAL REPORT 2020–2021

The C3.ai Digital Transformation Institute is a research consortium dedicated to accelerating the benefits of artificial intelligence for business, government, and society. The Institute engages the world's leading scientists to conduct research and train practitioners in the new science of digital transformation — operating at the intersection of artificial intelligence, machine learning, cloud computing, internet of things, big data analytics, organizational behavior, public policy, and ethics.

Table of Contents

THE NEW SCIENCE OF DIGITAL TRANSFORMATION

Thomas M. Siebel
Chairman and CEO, C3 AI

Photo credit:
Adam Lau/Berkeley Engineering

What is digital transformation? Some describe it as the power of digital technology applied to every aspect of the organization. Some refer to it as using digital technologies and advanced analytics for economic value, agility, and speed.

I find it more valuable to describe digital transformation through examples. This is partly because we are in the midst of massive disruption. Its implications are still evolving, and its impacts are still being understood. Each iteration — whether across companies or industries or even within a single organization — will bring new insights and layers to our understanding of digital transformation.

To be clear, digital transformation is not a series of generational changes in information technology or simply the migration of a company's processes, data, and information onto a digital platform. As industry analyst Brian Solis of Altimeter Group writes, "Investing in technology isn't the same as digital transformation."

To put digital transformation in context, let's take a step back and look at the previous waves of digital evolution. What we see is reminiscent of punctuated equilibrium — periods of stability followed by rapid change and disruption, resulting in new winners and losers. For those of us who have been in technology for several decades, we can retrace the waves of extraordinary productivity growth among organizations and governments. But we will see that the changes during past periods are vastly distinct from what we are experiencing today with digital transformation. Although digital transformation will bring about similar productivity benefits, these benefits will be achieved in a very different way.

"These new technologies will boost economic growth, promote inclusiveness, improve the environment, and extend the length and quality of human life."

What does the future of digital transformation hold? From where I sit, it's clear the benefit for business and society will be enormous — on the order of the Industrial Revolution. These new technologies will boost economic growth, promote inclusiveness, improve the environment, and extend the length and quality of human life.

Many of the benefits cannot even be conceived of today. The point is that digital transformation will have profound effects, but not necessarily the effects we can predict or even measure now. Clearly, the building blocks to enable digital transformation are available, robust, and accessible: cloud computing, big data, AI, and IoT.

– Tom Siebel, *Digital Transformation: Survive and Thrive in an Era of Mass Extinction*

LEADERSHIP

The C3.ai Digital Transformation Institute was established in March 2020 by C3 AI, Microsoft Corporation, the University of California, Berkeley, and the University of Illinois at Urbana-Champaign. Institute partners include Carnegie Mellon University, KTH Royal Institute of Technology, Lawrence Berkeley National Laboratory, Massachusetts Institute of Technology, National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, Princeton University, Stanford University, and University of Chicago. The Institute is jointly managed and hosted by the University of California, Berkeley and the University of Illinois at Urbana-Champaign.

Advisory Board

Thomas M. Siebel
Chairman and CEO, C3 AI

"The C3.ai Digital Transformation Institute is a consortium of leading scientists, researchers, innovators, and executives from academia and industry, joining forces to accelerate the social and economic benefits of digital transformation. We have the opportunity through public-private partnership to change the course of global challenges. I cannot imagine a more important use of AI. The probability of something good not coming out of this is zero – sparks are going to fly, change is going to happen, and the world is going to be a better place."

Eric Horvitz
Chief Scientific Officer, Microsoft

"We're happy to be part of a shared mission to accelerate research at the eminent research institutions that constitute the C3.ai Digital Transformation Institute. As we continue on this exciting private-public partnership, we're enthusiastic about aiming the broader goals of the Institute at acute challenges, such as mitigating the global pandemic, as well as on longer-term research that could help address other pressing challenges, such as energy, sustainability, and climate security. We want the top intellectual talent at leading universities to dive in with full force."

S. Shankar Sastry
Co-Director, C3.ai DTI
Thomas M. Siebel Professor of Computer Science

"We're dedicated to getting at the heart of the underlying science and technology of digital transformation. The underpinnings of this dynamic new science, which combines data from increasingly widespread IoT networks with analytics, stands at the intersection of several disciplines – including game theory, network science, mechanism design, reinforcement learning and inverse reinforcement learning, and fundamental concepts from cybersecurity and privacy. With tremendous interest from industry, it is critical that the technology be robust, resilient, and integrated with societal values that we cherish deeply."

R. Srikant
Co-Director, C3.ai DTI
Fredric G. and Elizabeth H. Nearing Endowed Professor of Electrical and Computer Engineering

"The Institute catalyzes research activities in mathematical, statistical, and computing research – including machine learning, artificial intelligence, the internet of things, and ethics and AI public policy. We aim to build a fundamental set of scientific advances, algorithms, designs, and business change management practices to establish the science of digital transformation of societal systems. C3.ai DTI will develop the tools to effect a digital transformation in basic and applied sciences and diverse industry sectors – and, throughout, to advance the social good."

Executive Committee

Jonathan Carter
Campus Lead, C3.ai DTI
Associate Lab Director for Computing Sciences
Lawrence Berkeley National Laboratory

Michael Franklin
Campus Lead, C3.ai DTI
Liew Family Chairman of Computer Science
University of Chicago

Karl H. Johansson
Campus Lead, C3.ai DTI
VR Distinguished Professor of Electrical Engineering and Computer Science
KTH Royal Institute of Technology

Christopher Manning
Campus Lead, C3.ai DTI
Thomas M. Siebel Professor of Machine Learning
Stanford University

Asuman Ozdaglar
Campus Lead, C3.ai DTI
MathWorks Professor of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

H. Vincent Poor
Campus Lead, C3.ai DTI
Michael Henry Strater University Professor
Princeton University

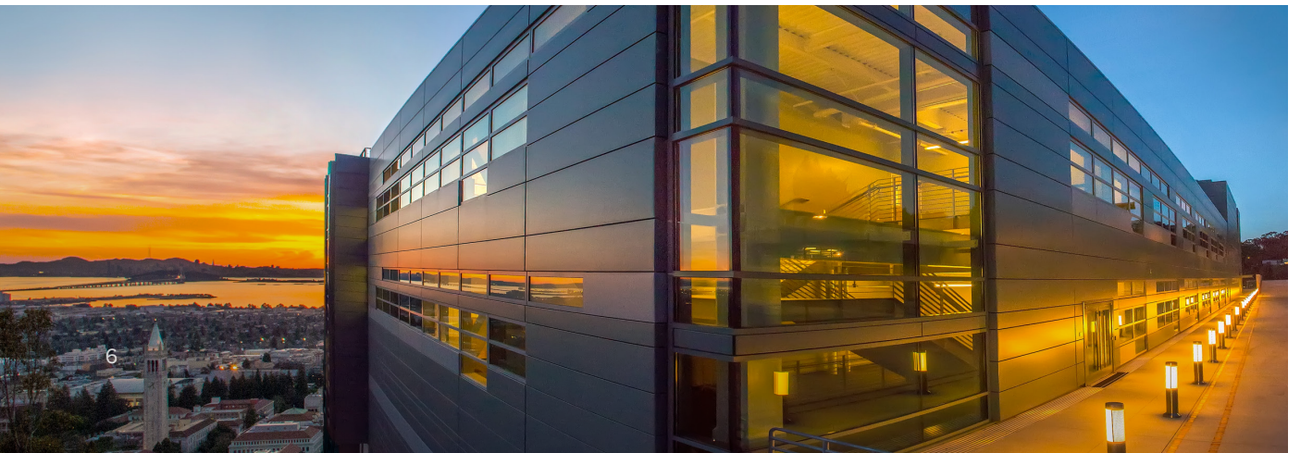
William Sanders
Campus Lead, C3.ai DTI
Dr. William D. and Nancy W. Strecker Dean, College of Engineering
Carnegie Mellon University

S. Shankar Sastry
Co-Director, C3.ai DTI
Thomas M. Siebel Professor of Computer Science
University of California, Berkeley

R. Srikant
Co-Director, C3.ai DTI
Fredric G. and Elizabeth H. Nearing Endowed Professor of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Costas Spanos
Co-Chief Scientist, C3.ai DTI
Professor of Electrical Engineering and Computer Sciences
University of California, Berkeley

Tandy Warnow
Co-Chief Scientist, C3.ai DTI
Grainger Distinguished Chair of Engineering
University of Illinois at Urbana-Champaign



PARTNER INSTITUTIONS

The C3.ai Digital Transformation Institute consortium includes members from academia, national laboratories, and industry. The Industry Partners program enables leading companies from around the world to engage with C3.ai DTI researchers and C3.ai DTI activities. Industry partners are also encouraged to engage with Institute researchers and collaborate on research projects and participate in the annual C3.ai DTI Research Symposium and attend workshops and colloquia.

Academic Partners

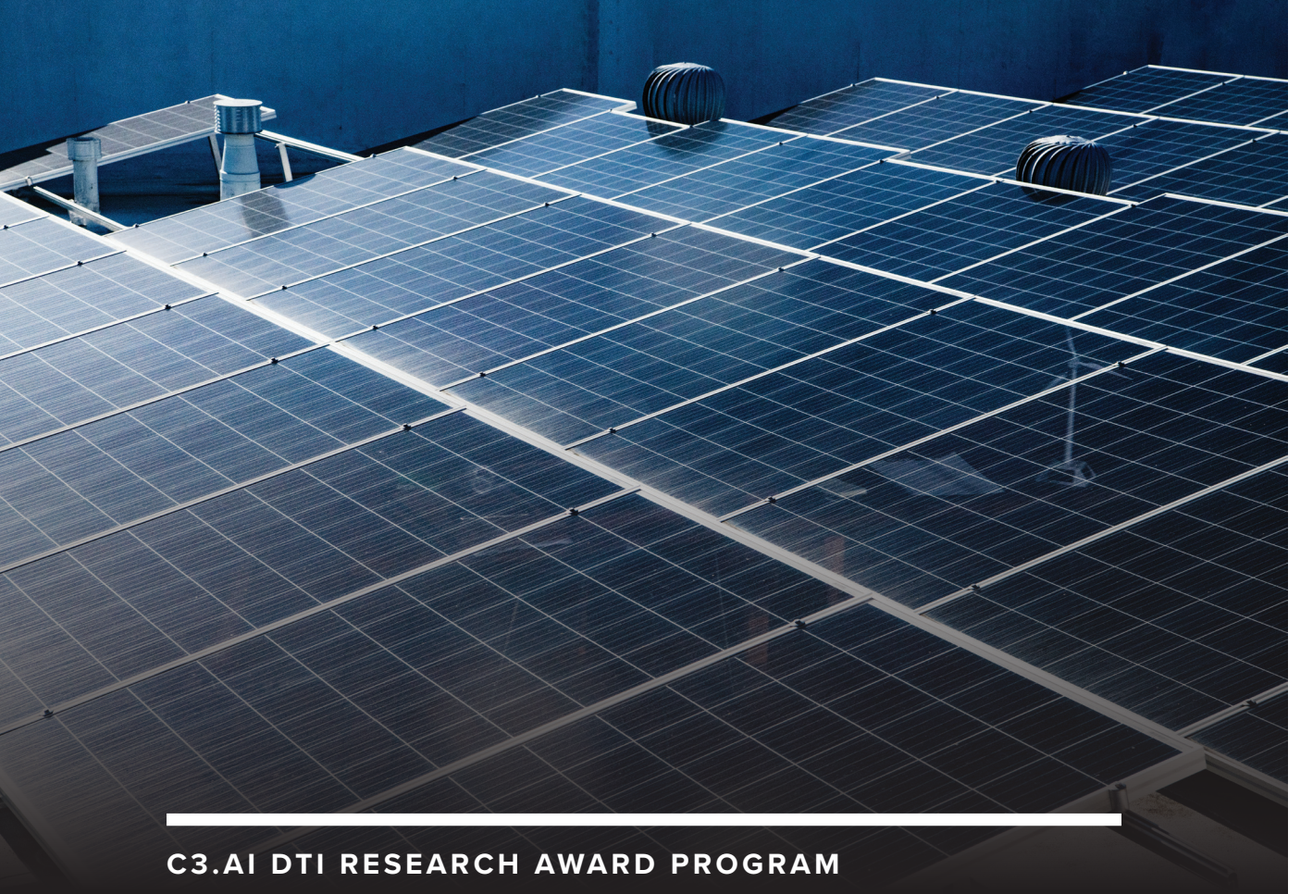
Carnegie Mellon University
KTH Vetenskap och Konst
Massachusetts Institute of Technology
Princeton University
Stanford University
Berkeley University of California
The University of Chicago
University of Illinois Urbana-Champaign

National Laboratories

Berkeley Lab
National Center for Supercomputing Applications, University of Illinois Urbana-Champaign

Industry Partners

AstraZeneca
Baker Hughes
C3.ai
Microsoft
Shell



C3.AI DTI RESEARCH AWARD PROGRAM

The C3.ai Digital Transformation Institute supports teams of the best researchers in the world to advance AI techniques for industrial, commercial, and public sector applications. This rich ecosystem will help address some of the most complex issues inherent in a massive societal digital transformation and build the foundation for a new Science of Digital Transformation. C3.ai DTI annually awards cash grants and access to computing resources for research projects based at consortium universities.

In spring 2020, in response to the worldwide challenges of the SARS-CoV-2 pandemic, the Institute's first call for proposals was for projects designed to mitigate the spread of COVID-19 and future pandemics — awarding 26 research proposals in six categories in June 2020 (see pages 9-17).

In spring 2021, to address another global threat, the Institute released its second call for proposals on digital transformation and AI for energy and climate security — inviting projects that apply AI/ML to sustainability initiatives, carbon sequestration, safe hydrocarbon production and transportation infrastructure, smart grid analytics, energy-efficient buildings, climate change modeling, and more. Awards will be announced in June 2021.

AI RESEARCH PROJECTS TO MITIGATE THE PANDEMIC

In its first call for proposals in 2020, the C3.ai Digital Transformation Institute selected 26 research proposals that advance digital transformation science to mitigate COVID-19 and future pandemics. A total of $5.4 million in funding and access to the C3 AI Suite and Microsoft Azure computing and storage were awarded in support of the projects.

AI FOR EPIDEMIOLOGY, SOCIAL GOOD, AND CLINICAL USE

Housing Precarity, Eviction, and Inequality in the Wake of COVID-19

Karen Chapple
University of California, Berkeley

Ensuring housing security is vital to mitigating the spread of COVID-19 and sustaining health, economic security, and family stability. This interdisciplinary project of academics, planners, and data scientists tracks housing evictions during and after the outbreak to better understand housing precarity and inform public policy regarding U.S. housing inequality. The team developed an innovative system for tracking real-time eviction filings after the outbreak, along with a housing precarity risk model using machine learning to better analyze and predict areas at highest risk of displacement.

Improving Fairness and Equity in COVID-19 Policy Applications of Machine Learning

Rayid Ghani
Carnegie Mellon University

When policy planning and resource allocation decisions are made using AI, there is a risk that they could result in inequitable and unfair outcomes for vulnerable populations. Disparate impacts of the pandemic on racial minorities and economically disadvantaged populations are already evident, and the risk that these disparities could worsen through applications of AI is substantial. This research focuses on developing bias detection/audit, reduction, and mitigation methods and tools to ensure that AI- and ML-based policy actions reduce the risk of inequitable outcomes for vulnerable populations.

Modeling the Impact of Social Determinants of Health on COVID-19 Transmission and Mortality to Understand Health Inequities

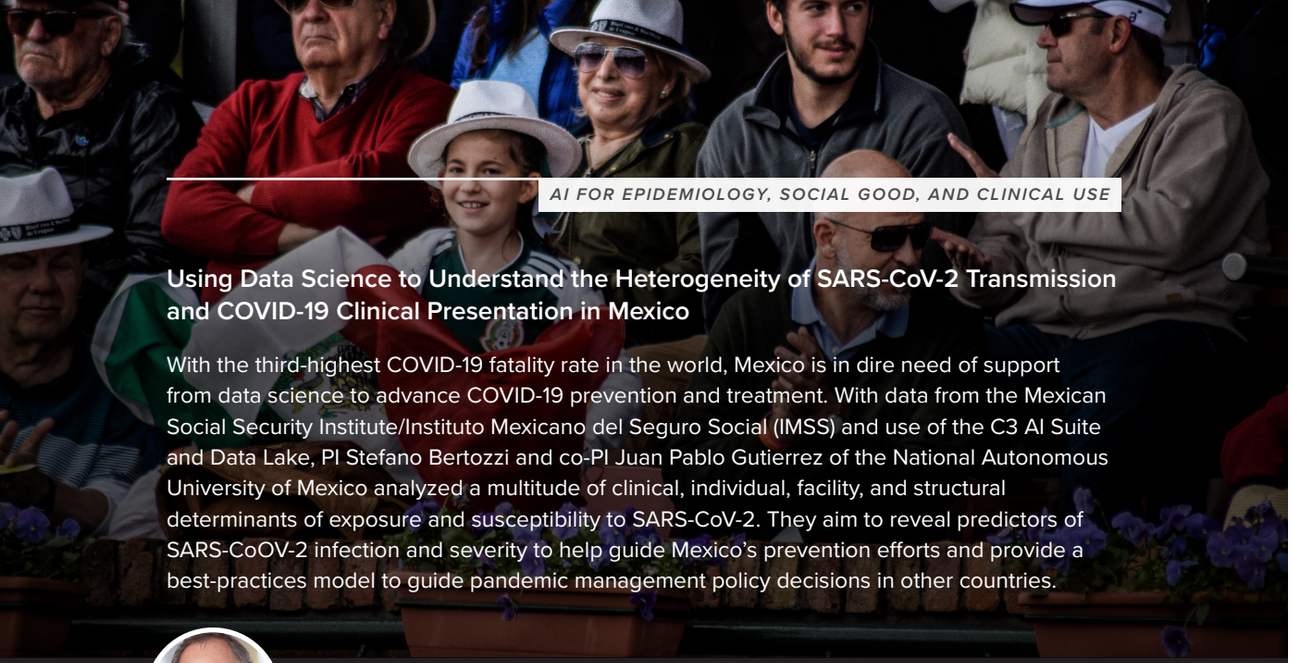
Anna Hotton
University of Chicago

Given disproportionate impact on populations most heavily affected by COVID-19, reducing morbidity and mortality requires increased focus on social determinants of health. Building upon Chicago's existing model of COVID-19 transmission, CityCOVID, this team applies machine learning methods to quantify the impact of social determinants of health on transmission — including financial hardship, housing instability, health care access, and incarceration — for a more realistic model to better inform policies around social distancing, testing, and vaccination scale-up.

Detection and Containment of Emerging Diseases Using AI

Alberto Sangiovanni-Vincentelli
University of California, Berkeley

When using AI to tackle the COVID-19 pandemic, distribution shifts from test data to training data proves particularly challenging. Shifts often degrade performance — critical when deploying machine-learning models in high-stakes decision-making scenarios, such as healthcare and autonomous driving. This research team is devising approaches to leverage the uncertainty from ensemble learners and domain randomization. Their theoretical and empirical results show that these approaches produce classifiers that are more robust against distribution shifts.



AI FOR EPIDEMIOLOGY, SOCIAL GOOD, AND CLINICAL USE

Using Data Science to Understand the Heterogeneity of SARS-CoV-2 Transmission and COVID-19 Clinical Presentation in Mexico

With the third-highest COVID-19 fatality rate in the world, Mexico is in dire need of support from data science to advance COVID-19 prevention and treatment. With data from the Mexican Social Security Institute/Instituto Mexicano del Seguro Social (IMSS) and use of the C3 AI Suite and Data Lake, PI Stefano Bertozzi and co-PI Juan Pablo Gutierrez of the National Autonomous University of Mexico analyzed a multitude of clinical, individual, facility, and structural determinants of exposure and susceptibility to SARS-CoV-2. They aim to reveal predictors of SARS-CoOV-2 infection and severity to help guide Mexico's prevention efforts and provide a best-practices model to guide pandemic management policy decisions in other countries.

Stefano Bertozzi
University of California, Berkeley

COVID-19 Medical Best-Practice Guidance System

Lui Sha
University of Illinois at Urbana-Champaign

Like GPS calculates routes in real-time, this group is developing real-time guidance for COVID-19 treatment to minimize risks of preventable medical errors when patient surges can overwhelm hospital staff. The web-based service, backed by a mathematically verifiable computational pathophysiology model, improves the efficacy of medical interventions. First under development, a simplified prototype for screening and management of Acute Respiratory Distress Syndrome (ARDS), the deadliest phase of COVID-19 pneumonia. The C3 AI Suite is used to provide early warning of patient adverse events.

See featured projects for two additional projects in this category, one led by Stefano Bertozzi, UC Berkeley, above, and one led by Stefana Parascho, Princeton, on opposite page.

MATHEMATICAL MODELING, CONTROL, AND LOGISTICS

Evolutionary Adaptations and Spreading Processes in Complex Networks

H. Vincent Poor
Princeton University

Many models for spreading processes in networks assume that the propagating object — such as a pathogen, in the context of infectious disease propagation, or, in the context of information propagation, a piece of information — is transferred across network nodes without modification. Yet like a child's game of telephone, information is most often modified as it gets passed along. Similarly, pathogens can mutate into new variants due to evolutionary phenomena. This team aims to reveal the impact of such adaptations on spreading processes in complex networks and their role in determining the threshold, probability, and final size of epidemics.

Impacts of COVID-19 Interventions: Health, Economics, and Inequality

Munther Dahleh
Massachusetts Institute of Technology

To reopen our universities, we need new technology and innovative practices to safeguard students against virus outbreak. This team is developing analytical methods for modeling and mitigating COVID-19 conditions based on student location and symptom data, collected via mobile apps. Aiming to strike a balance between containing the virus and keeping productive on-campus activities, the team leverages recent advances in system identification, reinforcement learning, and adaptive control to develop predictive methods to infer the hidden health states of individual students and develop algorithms to recommend optimal interventions for decision makers, such as testing and quarantine.

Impact of Mobility on Epidemic Spread: Lessons from NYC and India

Saurabh Amin
Massachusetts Institute of Technology

An examination of New York City's Metropolitan Transportation Authority (MTA) and public transport in India's eastern state of Odisha enabled this team to evaluate the impact of public transportation systems on COVID-19 spread. While higher levels of mobility correlate with an increase in cases, MTA use did not lead to additional case growth after April 2020. In contrast, Odisha data shows a strong correlation between case growth and the volume of incoming migrant workers from high-risk states after May. The team's optimization model allows health authorities to better allocate testing resources to worker populations based on risk levels at origin states and local district populations.

INTERDISCIPLINARY

Bringing Social Distancing to Light: Crowd Management Using AI and Interactive Floor Projection

Worldwide, governments are trying to identify ways to reopen public spaces and restart businesses and schools. Architecture can play a key role in the return to healthy public life by providing ways to control distances between people. With computational processing power and data accessibility, researchers are pursuing a multipronged approach to promote safe and efficient movement through public space. Their goal is to develop a computational tool that uses machine learning to predict people's movement and provides suggestions for adapting existing spaces through local physical interventions and systems based on light projections that offer direct real-time information about safe pedestrian trajectories and movement behavior.

Stefana Parascho
Princeton University

MATHEMATICAL MODELING, CONTROL, AND LOGISTICS

Dimitris Bertsimas
Massachusetts Institute of Technology

Toward Analytics-Based Clinical and Policy Decision Support to Respond to COVID-19 Pandemic

MIT researchers have developed a new epidemiological model of the disease's dynamics, a machine-learning model of mortality risk, and a resource allocation model — all published on covidanalytics.io. The team's automated, interpretable, and scalable decision-making systems are based on machine learning and artificial intelligence and support clinical practices and public policies responding to the COVID-19 pandemic. The CDC said of the effort, "Forecasts like this help us understand the most likely outcomes as well as best- and worst-case possibilities — and they can help policymakers make decisions that can lead us closer to those best-case outcomes." The work was also recognized with the INFORMS Pierskalla Award for the Best Paper in Healthcare in 2020.

Dynamic Resource Management in Response to Pandemics

Subhonmesh Bose
University of Illinois at Urbana-Champaign

With expertise in optimization and control, healthcare management, operations research, data analytics, and game theory, this multidisciplinary team aims to build a data analytic framework to optimize resource management for testing, prevention, and care. The research builds on work for the State of Illinois to mitigate the COVID-19 outbreak. Key outcomes include: a risk-aware dynamic equipment-and-workforce-allocation mechanism on the C3 AI Suite; comparing federalized resource allocation to states having to compete for resources; and flexible capacity-provisioning in medical supply chains and dynamic inventory management for critical goods, such as N-95 masks.

Algorithms and Software Tools for Testing and Control of COVID-19

Prashant Mehta
University of Illinois at Urbana-Champaign

This interdisciplinary project brings together epidemiologists, systems theorists, and data scientists to develop models, algorithms, and software tools to support state-level PCR (polymerase chain reaction) and serological testing efforts. The team is developing algorithms to assimilate real-time testing data into networked epidemiological models and mean-field-type control strategies to inform and evaluate the effect of social distancing and other control measures on disease progression. The team aims for improved epidemiological models that are more realistic — able to assimilate noisy data from ongoing population-level testing and including the effect of population-level feedback that may result from control measures.

Targeted Interventions in Networked and Multi-Risk SIR Models: How to Unlock the Economy During a Pandemic

Asuman Ozdaglar
Massachusetts Institute of Technology

What are optimal lockdown and testing policies for the containment of disease spread in networked environments? This team is developing lockdown models that consider epidemic and economic aspects simultaneously. Since individuals in the population have different risks and productivity levels, modeling must account for these differences. Also, interventions implemented by one region will affect neighboring regions, so coordination requires modeling how state interconnections and mobility patterns affect optimal responses. Results will show leaders and decision-makers how to optimally unlock the economy within a state and optimally coordinate efforts between states.

Metapopulation and Age-Structured Epidemic Models for the COVID-19 Pandemic

Zoi Rapti
University of Illinois at Urbana-Champaign

This research offers a comprehensive approach to the spatial dynamics of COVID-19 based on partial differential equation and metapopulation models — aiming to fill the gap between studies that detail disease dynamics but lack spatial dynamics and those that, while spatial in nature, do not account for the intricacies of COVID-19. Using the C3 AI Suite and techniques ranging from dynamic traffic assignment for metapopulation models to manifold learning techniques for model parameterization.

See featured projects for an additional project in this category, led by Dimitris Bertsimas, Massachusetts Institute of Technology, on opposite page.

VACCINE AND DRUG DISCOVERY

Ziv Bar-Joseph
Carnegie Mellon University

Effective Cocktail Treatments for SARS-CoV-2 Based on Modeling Lung Single-Cell Response Data

Fully modeling the impact of SARS-CoV-2 requires integrating several different types of molecular and cellular data. While SARS-CoV-2 is known to primarily impact cells via two viral entry factors, ACE2 and TMPRSS2, much less is known about virus activity within lung cells. To model the host response to viral infection and develop potential treatments, researchers are reconstructing pathways leading from virus proteins via their host interactions to regulators to the observed expression profiles in each cell. The research has demonstrated how coronavirus damages lung cells within hours; researchers have studied 18 FDA-approved drugs that could halt coronavirus infection earlier on.

VACCINE AND DRUG DISCOVERY

Machine Learning-Based Vaccine Design and HLA-Based Risk Prediction for Viral Infections

David Gifford
Massachusetts Institute of Technology

Based upon developments in predicting the presentation of viral antigens by Class I and Class II MHC complexes, this research develops new methods for vaccine design and viral disease severity prediction that can target any virus given its genome sequence. Methods to rapidly develop new vaccines and predict severity of evolved viruses based on genotype are particularly valuable given the potential escape of SARS-CoV-2 from vaccination by mutation. These methods will be implemented on the C3 AI Suite and other platforms and use cloud resources for vaccine combinatorial optimization that considers more than 39,000 peptides to select a compact set for vaccine formulation.

Scoring Drugs: Small Molecule Drug Discovery for COVID-19 using Physics-Inspired Machine Learning

Teresa Head-Gordon
University of California, Berkeley

Molecular modeling is combating the current global pandemic through the traditional process of drug discovery, but slow turnaround times for identifying leads for antiviral drugs, analyzing structural effects of genetic variation in an evolving virus, and targeting relevant virus-host protein interactions remains a significant limitation in an acute crisis. This team uses machine learning techniques inspired by physics to speed the discovery of small molecules that could bind and disable the SARS-CoV-2 virus, leading to future drugs to treat the disease.

IMAGING/COMPUTER VISION

Maryellen Giger
University of Chicago

Medical Imaging Domain-Expertise Machine Learning for the Interrogation of COVID-19

The COVID-19 pandemic has revealed the pressing public health need for advanced computational techniques to improve interpretation of medical images for surveillance, detection, and triaging of COVID-19 medical images, for differential diagnosis of COVID-19 patients, and for prognosis. Thoracic imaging, including chest radiography and computed tomography, is limited, in part due to qualitative interpretation of the images. The team is developing machine-intelligence methods to help interrogate medical images from COVID-19 patients with the aim of demonstrating cascade-based deep transfer learning between different thoracic disease states and developing a clinical tool to better triage COVID-19 patients.

Data-Driven, High-Dimensional Design for Trustworthy Drug Discovery

Jennifer Listgarten
University of California, Berkeley

Data-driven design is making headway into a number of application areas, including protein, small-molecule, and materials engineering. This research draws upon techniques such as reinforcement learning, robust uncertainty estimation and probabilistic modeling to develop new and trustworthy methods for therapeutic drug discovery for COVID-19. The design goal is to construct an object with desired properties, such as a protein therapeutic that binds tightly to its target.

See featured projects for an additional project in this category, led by Ziv Bar-Joseph, Carnegie Mellon University, on page 13.

IMAGING/COMPUTER VISION

Triaging of COVID-19 Patients from Audio-Visual Cues

Narendra Ahuja
University of Illinois at Urbana-Champaign

This team is developing audiovisual tools to reproduce common physical exam findings in COVID-19 patients to expand patient stability assessment methods to, ultimately, conserve hospital beds. Tools could be used to predict clinical decompensation from patient videos captured using consumer-grade smartphones. The team is developing AI/ML algorithms for predicting impending deterioration from health-relevant audiovisual features and to provide explanations of clinical details within electronic health records. The techniques and algorithms are likely adaptable to other high-risk patient populations and emerging platforms, such as telemedicine.

Nancy Amato
University of Illinois at Urbana-Champaign

Tracking the Few and Far Between: Computational Strategies to Speed the Discovery of Low-Frequency Genomic Variation in COVID-19

The vast majority of COVID-19 genomic research focuses on a high-level view of SARS-CoV-2 diversity, overlooking viral population diversity within each COVID-19-positive patient. An individual's SARS-CoV-2 viral load may exceed hundreds of thousands of copies, yet available genomic databases only contain a single consensus version of this diverse population, discarding low-frequency mutations. The CoVariants team develops novel computational approaches to recover discarded variants and permit rapid characterization of within-host diversity of SARS-CoV-2 across tens of thousands of samples. The team's new parallelization strategies and approximate statistical methods can reduce computational requirements of a commonly used approach by up to 400 percent, while preserving 100 percent of low-frequency genomic diversity. These improved methods can provide insight into the biological underpinnings of SARS-CoV-2 transmissibility and severity compared to other coronaviruses.

COMPUTATIONAL BIOLOGY

DISTRIBUTED COMPUTING

Towards AI for Healthcare with Applications to the COVID-19 Pandemic

Artificial Intelligence is among the most powerful tools for addressing early detection of diseases, triage, treatment planning, and patient management, among many other pressing healthcare problems. So what will it take to build effective machine-learning systems for healthcare? This research endeavors to answer this question, using emerging technical advances in federated learning, modeling, evaluation, privacy, and trustworthiness, and bringing these tools to bear to aid in analyzing medical images from COVID-19 patients.

Sanmi Koyejo
University of Illinois at Urbana-Champaign

Machine Learning Support for Emergency Triage of Pulmonary Collapse in COVID-19

Sendhil Mullainathan
University of Chicago

Working with frontline doctors treating COVID-19, this team aims to produce an algorithm to predict pulmonary collapse, and thereby help physicians make better triage decisions. With access to four million chest X-rays linked to physiological markers of pulmonary collapse (including ARDS, or acute respiratory distress syndrome, the "final common pathway" for many infections) and mortality from linked Social Security data, the modern machine learning toolkit could be deployed for COVID-19 and more broadly as a general prediction toolkit for pulmonary collapse.

See featured projects for an additional project in this category, led by Maryellen Giger, University of Chicago, on page 14.

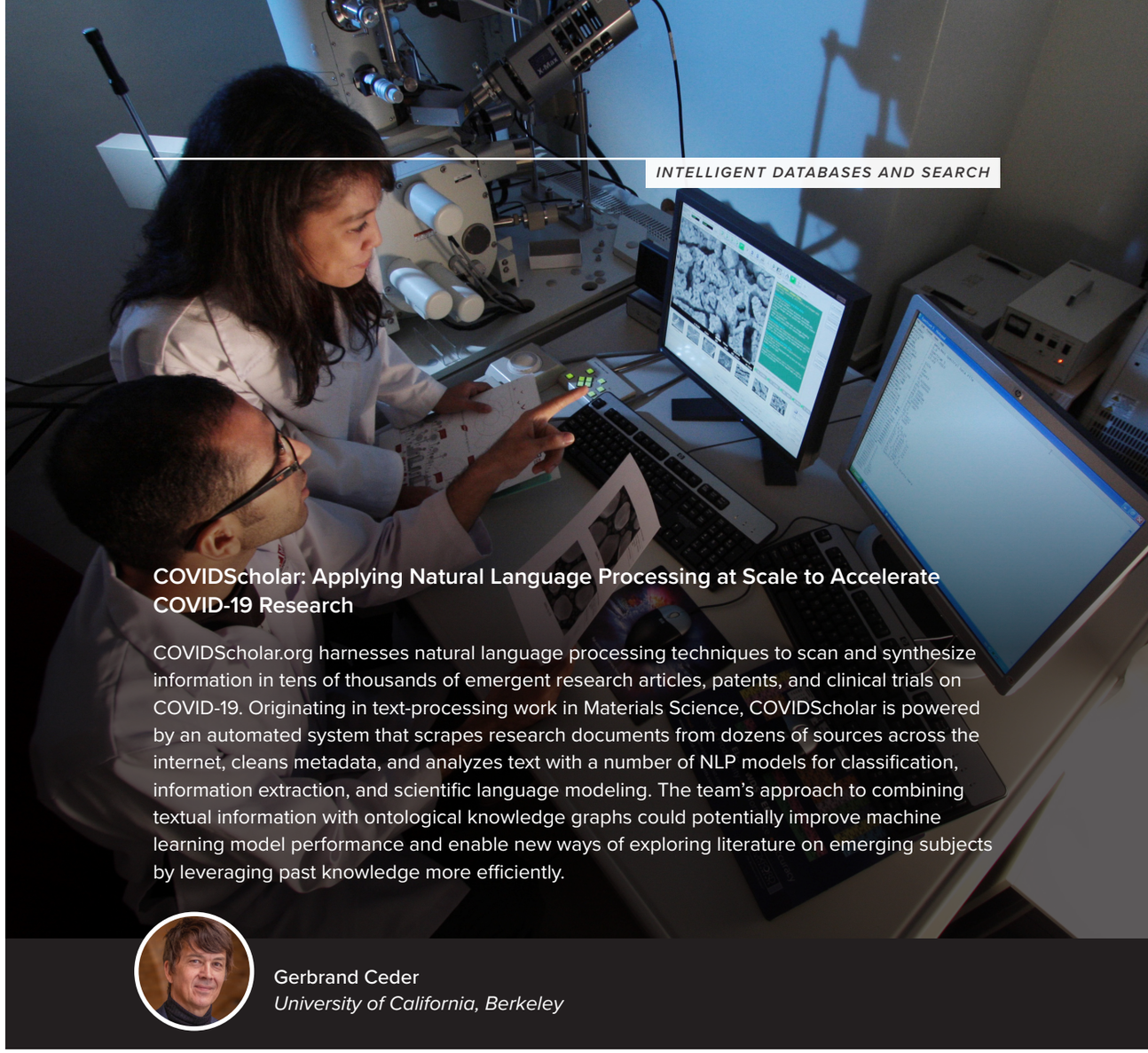
COMPUTATIONAL BIOLOGY

AI-Enabled Deep Mutational Scanning of Interaction Between SARS-CoV-2 Spike Protein S and Human ACE2 Receptor

Diwakar Shukla
University of Illinois at Urbana-Champaign

This research team is investigating the relationship between the SARS-CoV-2 Spike Protein S and the ACE2 receptor. Virus mutations can bind to this receptor, strengthening their ability to spread infections. By examining the sequence, structure, and function in this key relationship, researchers aim to engineer the ACE2 receptor and create a decoy receptor that could be injected into the body to neutralize the SARS-CoV2 virus. "We want to design new proteins that can neutralize coronavirus – also, if there are mutations, how can we change our design to target that virus, and how can we make that cycle really fast," says Shukla.

See featured projects for an additional project in this category, led by Nancy Amato, University of Illinois, at Urbana-Champaign, on page 15.



INTELLIGENT DATABASES AND SEARCH

COVIDScholar: Applying Natural Language Processing at Scale to Accelerate COVID-19 Research

COVIDScholar.org harnesses natural language processing techniques to scan and synthesize information in tens of thousands of emergent research articles, patents, and clinical trials on COVID-19. Originating in text-processing work in Materials Science, COVIDScholar is powered by an automated system that scrapes research documents from dozens of sources across the internet, cleans metadata, and analyzes text with a number of NLP models for classification, information extraction, and scientific language modeling. The team's approach to combining textual information with ontological knowledge graphs could potentially improve machine learning model performance and enable new ways of exploring literature on emerging subjects by leveraging past knowledge more efficiently.

Gerbrand Ceder
University of California, Berkeley

INTELLIGENT DATABASES AND SEARCH

See featured projects for a project in this category, led by Gerbrand Ceder, University of California, Berkeley, above.

DISTRIBUTED COMPUTING

See featured projects for a project in this category, led by Sanmi Koyejo, University of Illinois at Urbana-Champaign, on opposite page.

FRONTIERS OF DIGITAL TRANSFORMATION SCIENCE

INAUGURAL C3.AI DIGITAL TRANSFORMATION INSTITUTE ANNUAL RESEARCH SYMPOSIUM 2021

The annual C3.ai DTI Research Symposium is designed to bring together from around the globe leaders and students of the new science of digital transformation. The inaugural symposium in January 2021 showcased leading research and applications of artificial intelligence and machine learning that are advancing the science, including the use of AI and ML to mitigate the COVID-19 pandemic.

Program | Day 1: January 20, 2021

Welcome:

Shankar Sastry, C3.ai DTI Co-Director, University of California, Berkeley and R. Srikant, C3.ai DTI Co-Director, University of Illinois at Urbana-Champaign

Opening Remarks:

Thomas Siebel, Chairman and CEO, C3 AI

Keynote:

"Modeling the Spread and Mitigation of COVID-19 in a Large Public University," Nigel Goldenfeld, University of Illinois at Urbana-Champaign

"Using Data Science to Understand the Heterogeneity of SARS-COV-2 Transmission and COVID-19 Clinical Presentation in Mexico," Stefano Bertozzi, University of California, Berkeley

"Effective Cocktail Treatments for SARS-CoV-2 Based on Modeling Lung Single Cell Response Data," Ziv Bar-Joseph, Carnegie Mellon University

"Secure Federated Learning for Clinical Informatics with Applications to the COVID-19 Pandemic," Dakshita Khurana, University of Illinois at Urbana-Champaign

"Adding Audio-Visual Cues to Signs and Symptoms for Triaging Suspected or Diagnosed COVID-19 Patients," Narendra Ahuja, University of Illinois at Urbana-Champaign

"Algorithms and Software Tools for Testing and Control of COVID-19," Prashant Mehta, University of Illinois at Urbana-Champaign

"Targeted Interventions in Networked and Multi-Risk SIR Models: How to Unlock the Economy During a Pandemic," Francesca Parise, Cornell University

"Spatial Modeling of Covid-19: Optimizing PDE and Metapopulation Models for Prediction and Spread Mitigation," Zoi Rapti, University of Illinois at Urbana-Champaign

"Detection and Containment of Emerging Diseases Using AI Technique," Alberto Sangiovanni-Vincentelli, University of California, Berkeley

"AI-enabled Deep Mutational Scanning of Interaction between SARS-CoV-2 Spike Protein S and Human ACE2 Receptor," Diwakar Shukla, University of Illinois at Urbana-Champaign

Program | Day Two: January 21, 2021

Keynote:

"AI Partners in the Classroom: Challenges and Opportunities," Martha Palmer, University of Colorado at Boulder

"Data-Driven, High-Dimensional Design for Trustworthy Drug Discovery," Sergey Levine, University of California, Berkeley

"Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID," Maryellen Giger, University of Chicago

"Modeling the Impact of Social Determinants of Health on COVID-19 Transmission and Mortality to Understand Health Inequities," Anna Hotton, University of Chicago

"Toward Analytics-Based Clinical and Policy Decision Support to Respond to the COVID-19 Pandemic," Dimitris Bertsimas, Massachusetts Institute of Technology

"Improving Fairness and Equity in COVID-19 Policy Applications of Machine Learning," Kit Rodolfa, Carnegie Mellon University

"Model Choice and Structure: Connections to Prediction, Policy, and Values," Ben Schaffer, Princeton University

"Housing Precarity, Eviction, and Inequality in the Wake of COVID-19," Karen Chapple, University of California, Berkeley

"Dynamic Resource Management in Response to Pandemics," Ujjal Mukherjee, University of Illinois at Urbana-Champaign

"Reinforcement Learning to Safeguard Schools and Universities Against the COVID-19 Outbreak," Munther Dahleh, Massachusetts Institute of Technology

"The extensive modeling behind UIUC's much-lauded SHIELD program that enabled the university to re-open sooner and remain open last year was unique in that it was predicated on the physics of aerosol transmission, even before the CDC was identifying aerosols as the primary means of transmission."

Nigel Goldenfeld, University of Illinois at Urbana-Champaign

KEYNOTE SPEAKER

Martha Palmer, University of Colorado at Boulder

KEYNOTE SPEAKER

"For intelligent computers to work side-by-side with students, researchers need to advance the fundamental science of how machines process human language, gestures, and emotions — and overcome the unique algorithmic and ethical challenges of analyzing and generating spoken language and detecting gestures, facial expressions, and emotions."

WORKSHOPS ON DIGITAL TRANSFORMATION SCIENCE

The Workshops on Digital Transformation Science are deep dives into foundational topics in digital transformation. These multi-day events are intended for researchers, practitioners, policymakers, and others interested in gaining insights and understanding of topics and trends from leading experts and scientists. Find all workshop videos on the C3.ai DTI YouTube channel at: *YouTube.com/C3DigitalTransformationInstitute*.

SEPTEMBER 8–11, 2020

Epidemics, Opinion, and (Mis)Information: Analytic Foundations of Dynamics over Networks

Organizers:

Devavrat Shah, Massachusetts Institute of Technology, and Lei Ying, University of Michigan

Abstract:

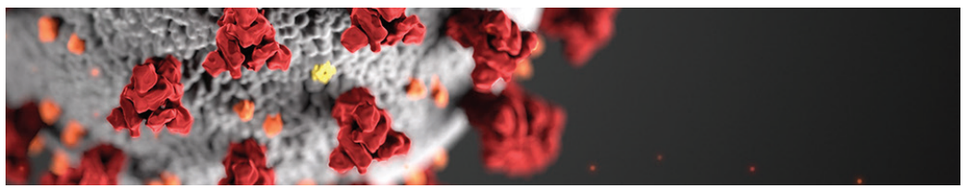
Epidemics spread through networks of human contact. Such networks have been studied for close to a century, and understanding them has never been more important as we face the COVID-19 pandemic. Opinions are formed similarly through social interactions, in person or through the print and broadcast media, and increasingly over social media, where "information dynamics" can quickly influence perceptions and shape behavior. It has become increasingly urgent to understand how the spread of information through social media platforms, driven by bots, bad actors, and unsuspecting users, are creating networks that can result in extreme polarization, information echo chambers, and the proliferation of "fake news."

Tutorials:

Experts from a range of disciplines examined three types of networks — distributed databases, distributed computational frameworks, and more generally distributed autonomous agents — to build common foundations about such networks and decipher their differences.

Speakers:

Noah Friedkin, University of California, Santa Barbara; Matthew O. Jackson, Stanford University; Jon Kleinberg, Cornell University; Naomi Ehrich Leonard, Princeton University; Laurent Massoulie, Inria, Microsoft Research; Elchanan Mossel, Massachusetts Institute of Technology; H. Vincent Poor, Princeton University; Weina Wang, Carnegie Mellon University; Tauhid Zaman, Yale University



OCTOBER 19–21 & 23, 2020

Analytical Foundations of Deep Learning: Interpretability and Performance Guarantees

Organizers:

Yi Ma, University of California, Berkeley, and René Vidal, Johns Hopkins University

Abstract:

We are on the verge of a deep learning revolution that is leading to many disruptive technologies: from automatic speech recognition systems such as Apple Siri, to automated supermarkets such as Amazon Go, to autonomous vehicles such as Google Car. As we increasingly employ deep learning in our daily lives to support important decisions, it becomes critical to understand the predictions made by deep neural networks.

Tutorials:

This workshop focused on "Principled design and interpretability" for Day 1 and "Guaranteed robustness and fairness" for Day 2. The first tutorial provided a mathematical justification for properties of conventional deep networks, such as global optimality, invariance, and stability of the learned representations. The second tutorial covered more recent developments on graph neural networks that are applicable to broader family of data structures.

Speakers:

Peter Bartlett, University of California, Berkeley; Tom Goldstein, University of Maryland; Gitta Kutyniok, Ludwig-Maximilians Universität München; Yi Ma, University of California, Berkeley; Alejandro Ribeiro, University of Pennsylvania; Guillermo Sapiro, Duke University; René Vidal, Johns Hopkins University; Soledad Villar, Johns Hopkins University; Max Welling, University of Amsterdam; Bin Yu, University of California, Berkeley

DECEMBER 7–8, 2020

Safe Autonomy: Learning, Verification, and Trusted Operation of Autonomous Systems

Organizers:

Geir Dullerud, University of Illinois at Urbana-Champaign, and Claire Tomlin, University of California, Berkeley

Abstract:

Advances in machine learning have accelerated the introduction of autonomy in our everyday lives. However, ensuring that these autonomous systems act as intended is an immense challenge. Today, when self-driving vehicles or collaborative robots operate in real-world uncertain environments, it is impossible to guarantee safety at all times. In this workshop, we explore the scope of safe autonomy, present and identify the challenges, and examine current research developments to help us move towards a solution.

Tutorials:

Control and robotics, AI and machine learning, formal methods, and human-robot interaction, and their applications to the domains of ground, air, and space vehicles, also medical robotics.

DECEMBER 7–8, 2020 | CONTINUED

Speakers:

Pieter Abbeel, University of California, Berkeley; Lars Blackmore, SpaceX; J-P Clarke, University of Texas at Austin; Anca Dragan, University of California, Berkeley; Katie Driggs-Campbell, University of Illinois at Urbana-Champaign; Hadas Kress-Gazit, Cornell University; Sayan Mitra, University of Illinois at Urbana-Champaign; Sandeep Neema, Defense Advanced Research Projects Agency; George Pappas, University of Pennsylvania; Daniela Rus, Massachusetts Institute of Technology; Dawn Tilbury, National Science Foundation, University of Michigan; Keenan Wyrobek, Zipline

COLLOQUIA ON DIGITAL TRANSFORMATION SCIENCE

The Colloquium on Digital Transformation is a series of weekly online talks on how artificial intelligence, machine learning, and big data can lead to scientific breakthroughs with large-scale societal benefit. DTI presented 22 talks in the summer and fall of 2020 to a total audience of 1,868 participants in 31 countries over six continents.

JULY 9, 2020

"Translating AI Research in Breast Cancer Imaging to COVID-19"
Maryellen Giger, *University of Chicago*

JULY 16, 2020

"Using AI Techniques for Detection and Containment of COVID-19 and Emerging Diseases"
Alberto Sangiovanni-Vincentelli
University of California, Berkeley

JULY 23, 2020

"Can Targeted Closures Reduce Economic Loss and Control COVID-19 Spread?"
John Birge
University of Chicago Booth School of Business

JULY 30, 2020

"Networked Epidemiology Models for COVID-19 Analysis and Control"
Carolyn Beck
University of Illinois at Urbana-Champaign

AUGUST 6, 2020

"Optimal Targeted Lockdowns for COVID-19 in a Multi-Group SIR Model"
Daron Acemoglu
Massachusetts Institute of Technology

AUGUST 13, 2020

"Predictive and Prescriptive Analytics for the COVID-19 Pandemic"
Dimitris Bertsimas
MIT Sloan School of Management

AUGUST 20, 2020

"Lessons from COVID-19: Efficiency vs. Resilience"
Moshe Y. Vardi, *Rice University*

AUGUST 27, 2020

"Targeted Dynamic Interventions in Networked Epidemic Models"
Asuman Ozdaglar
Schwarzman College of Computing, MIT

SEPTEMBER 3, 2020

"Metapopulation and Age-Structured Epidemic Models for the COVID-19 Pandemic"
Zoi Rapti
University of Illinois at Urbana-Champaign

SEPTEMBER 10, 2020

"Evolutionary Adaptations and Spreading Processes in Complex Networks"
H. Vincent Poor, *Princeton University*

SEPTEMBER 17, 2020

"Impact of Mobility on Epidemic Spread: Some Lessons from NYC and India"
Saurabh Amin
Massachusetts Institute of Technology

SEPTEMBER 24, 2020

"Towards AI for Healthcare with Applications to the COVID-19 Pandemic"
Sanmi Koyejo
University of Illinois at Urbana-Champaign

OCTOBER 1, 2020

"Improving Fairness and Equity in Policy Applications of Machine Learning"
Rayid Ghani, *Carnegie Mellon University*

OCTOBER 8, 2020

"Solving 'Prediction Problems' in Health, from Heart Attacks to COVID-19"
Ziad Obermeyer
University of California, Berkeley

OCTOBER 15, 2020

"COVIDScholar: Applying Natural Language Processing at Scale to Accelerate COVID-19 Research"
Gerbrand Ceder
University of California, Berkeley

OCTOBER 22, 2020

"Machine Learning-Based Design of Proteins, Small Molecules, and Beyond"
Jennifer Listgarten
University of California, Berkeley

OCTOBER 29, 2020

"Reliable Predictions? Counterfactual Predictions? Equitable Treatment? Some Recent Progress in Predictive Inference"
Emmanuel Candès, *Stanford University*

NOVEMBER 5, 2020

"Reconstructing SARS-COV-2 Response Pathways"
Ziv Bar-Joseph, *Carnegie Mellon University*

NOVEMBER 12, 2020

"Mathematics of Deep Learning"
René Vidal, *Johns Hopkins University*

NOVEMBER 19, 2020

"Tracking the Few and Far Between: Computational Strategies to Speed the Discovery of Low-Frequency Genomic Variation in COVID-19"
Nancy Amato
University of Illinois at Urbana-Champaign

DECEMBER 3, 2020

"Stochastic Optimization of Inventory at Large-Scale Supply Chains"
Mehdi Maasoumy, *C3 AI*

DECEMBER 10, 2020

"Housing Precarity, Eviction, and Inequality in the Wake of COVID-19"
Karen Chapple, *University of California, Berkeley*

A NEW VISION FOR ENERGY AND CLIMATE SECURITY

PREVIEW 2021-2022

How can the science of digital transformation create new energy systems and ensure climate security for all? This is what the C3.ai Digital Transformation Institute has challenged scholars, researchers, and software developers to discover with its second call for proposals in spring 2021. This call aims to advance AI at scale for safer, cleaner, lower-cost, and more reliable energy for a lower-carbon, higher-efficiency future.

> *"Mitigating the impact of global energy generation will require a massive transformation of the world's energy infrastructure."*
>
> *Thomas M. Siebel*
> *Chairman and CEO, C3 AI*

Funded research projects will tackle energy challenges from multiple dimensions. Final projects to be awarded will be announced in June 2021.

Sustainability
- Applying AI, machine learning, and advanced analytic techniques to support sustainability initiatives for energy consumption and greenhouse gas emissions

AI for Carbon Sequestration
- Applying AI/ML techniques to increase the scale and reduce the cost of carbon sequestration

AI for Leaks and Emissions Detection
- Applying advanced AI/ML techniques for large-scale emissions detection, facility-level data reconciliation and gap analysis for emissions sensors, prediction of emissions risk, and analysis and optimization of flaring intensity across upstream and downstream operations

Safe Hydrocarbon Production and Transportation Infrastructure
- Applying analytic and AI/ML modeling techniques to increase the safety and reduce emissions from oil and gas extraction, petrochemical production, and hydrocarbon transportation

AI for Advanced Energy and Carbon Markets
- Enabling dynamic, automated, and real-time pricing of energy generation sources

Smart Grid Analytics
- Applying AI and other analytic approaches to improve the efficiency and effectiveness of grid transmission and distribution operations

Cybersecurity of Power and Energy Infrastructure
- Leveraging AI/ML techniques to improve the cybersecurity of our critical power and energy assets, as well as smart connected factories and homes

Distributed Energy Resource Management
- Applying AI to increase the penetration and use of distributed renewables

AI for Energy-Efficient Buildings and Factories
- Leveraging AI techniques for advanced building and factory control to improve energy efficiency

AI for Improved Natural Catastrophe Risk Assessment
- Applying AI to improve modeling of natural catastrophe risks from future weather-related events (e.g., tropical storms, wildfires, and floods)

Resilient Energy Systems
- Addressing how the use of AI/ML techniques and markets for energy and carbon introduce new vulnerabilities

AI for Improved Climate Change Modeling
- Use of AI/ML to address climate change modeling and adaptation

For all updates on the C3.ai Digital Transformation Institute, see the website at C3aiDTI.ai.

CONTACT

C3.ai Digital Transformation Institute @ Berkeley
University of California, Berkeley
750 Sutardja Dai Hall, MC 1764
Berkeley, California 94720-1764

C3.ai Digital Transformation Institute @ Illinois
University of Illinois at Urbana-Champaign
1205 W. Clark Street, MC-257, Room 1008
Urbana, Illinois 61801

Inquiries and more information

For general inquiries:
info@c3dti.ai

Sign up for our newsletter
at our website: C3DTI.ai

C3DTI.ai

@C3DTI /C3DTI

/company/c3-ai-digital-transformation-institute /c/C3DigitalTransformationInstitute