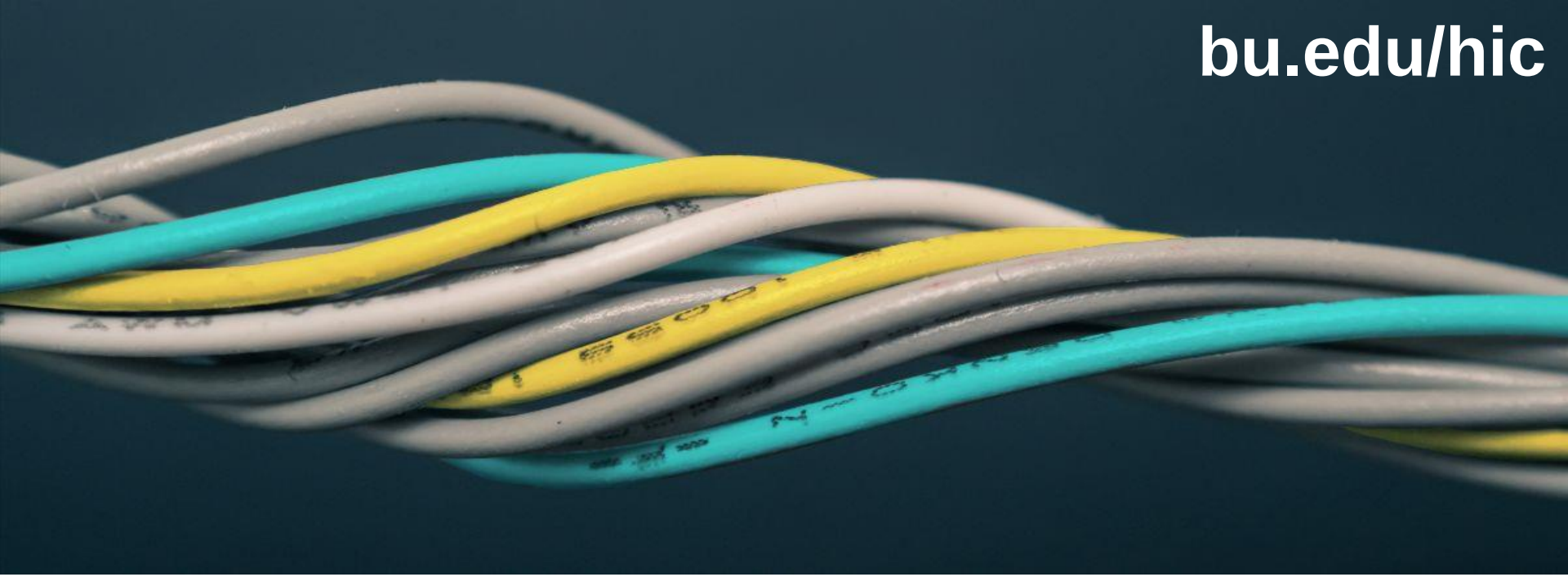


Come Together, Right Now

Convergence and Collaboration in Cloud Computing, Data Security & Artificial Intelligence

September 30, 2020

Eric Kolaczyk | Orran Krieger | Kate Saenko | Mayank Varia

Boston University Hariri Institute for Computing

BOSTON UNIVERSITY

Convergence and Collaboration in Cloud Computing, Data Security & Artificial Intelligence - Agenda

4:00 - 4:05	Introduction and Welcome, Gloria Waters
4:05 - 4:15	Hariri Institute 101, Eric Kolaczyk
4:15 - 4:25	Cloud, Orran Krieger
4:25 - 4:35	Privacy/Security, Mayank Varia
4:35 - 4:45	AI, Kate Saenko
4:45 - 5:25	Panel Discussion and Q&A
5:25 - 5:30	Closing Remarks, Eric Kolaczyk

HARIRI INSTITUTE 101

Eric Kolaczyk

ABOUT THE INSTITUTE

The Hariri Institute for Computing at Boston University is dedicated to leading integrated initiatives in research and technology development, targeting a broad set of disciplines at the nexus of the computational and data sciences.

The Institute also serves as a ***computational lens*** — into the impact and potential inherent in Boston University's computational and data-driven investments.

Mechanisms & Resources

The Institute leverages a diverse set of mechanisms and resources:

- thematic research centers and initiatives
- focused research programs
- software and data science development capacities
- lab/office space
- state-of-the-art conferencing facilities
- a spectrum of staff capabilities.

(And, these days, Zoom!)

Powering the Institute

- Supporting, nurturing fledgling initiatives
- Amplifying faculty research
- Grant strategy, submission, management
- Event planning, promotion
- Financial and legal operations, procurement, and payroll

Natalie McKenna *(nataliec@bu.edu)*
Associate Director, Finance & Administration

Emily Johnson *(emilypj@bu.edu)*
Finance & Operations Manager

Erica Seymourian *(eseymou1@bu.edu)*
Administrative Coordinator

Katherine D'Angelo *(ktd@bu.edu)*
Program & Events Manager

Korinne Dizon *(kdizon@bu.edu)*
Finance Manager

Reach out to us!

A Closer Look

- Centers, Initiatives, and Labs
- Focused Research Programs (FRPs)
- *"Did you know you could…?"* Series
- Hariri Institute Distinguished Speaker Series

(Note: All supported virtually for now, and likely hybrid going forward.)

Hariri Institute Centers, Initiatives, and Labs

Focused Research Programs

- Medium/large-group research efforts around year-long themes.
- Aligned with BU strategic priorities and/or emerging opportunities.
- Umbrellas over several “verticals” -- working-groups at the heart.
- Organization/goals inspired by specific funding mechanisms.
- Structured around a ‘package’ of HIC-facilitated support.

FRPs for AY20-21:

1. Sci. Machine Learning for Chemistry & Materials Science
2. AI and Medicine -- Bias and Underserved Populations

“Did You Know You Could …?” Series

- **WHO**: For anyone! Organized by the Graduate Student Fellows
- **WHAT**: A “brown bag” lunch series
- **WHEN**: Monthly (eventually biweekly?)
- **HOW**: 1 hr with Social/Lightening/Discussion format
- **WHY**: A way to get people from all walks @ BU to meet, mingle, and get exposed quickly and easily to things they did not know they would like to know.

Hariri Institute Distinguished Speaker Series

- **WHO**: For anyone! Organized by the Junior Faculty Fellows
- **WHAT**: Cluster of broadly accessible talks, plus panel discussion, by up-and-coming movers/shakers in computing and data science.
- **WHEN**: Once / semester
- **HOW**: 3 talks over ~2 weeks, followed group panel discussion
- **WHY**: Shine a spotlight on major challenges -- and emerging solutions -- in computing-enabled, data-driven topics of broad societal interest and impact.

Areas of Core Strength in Computing

From the perspective of computing, the Institute has core strengths in:

- Cloud computing
- Cybersecurity & privacy
- Artificial Intelligence

Emphasis is on both development of core areas and research convergence around domains and applications.

Today's event is centered on activity, impact, and potential around that core.

CLOUD PRESENTATION

Orran Krieger

Boston University Hariri Institute for Computing

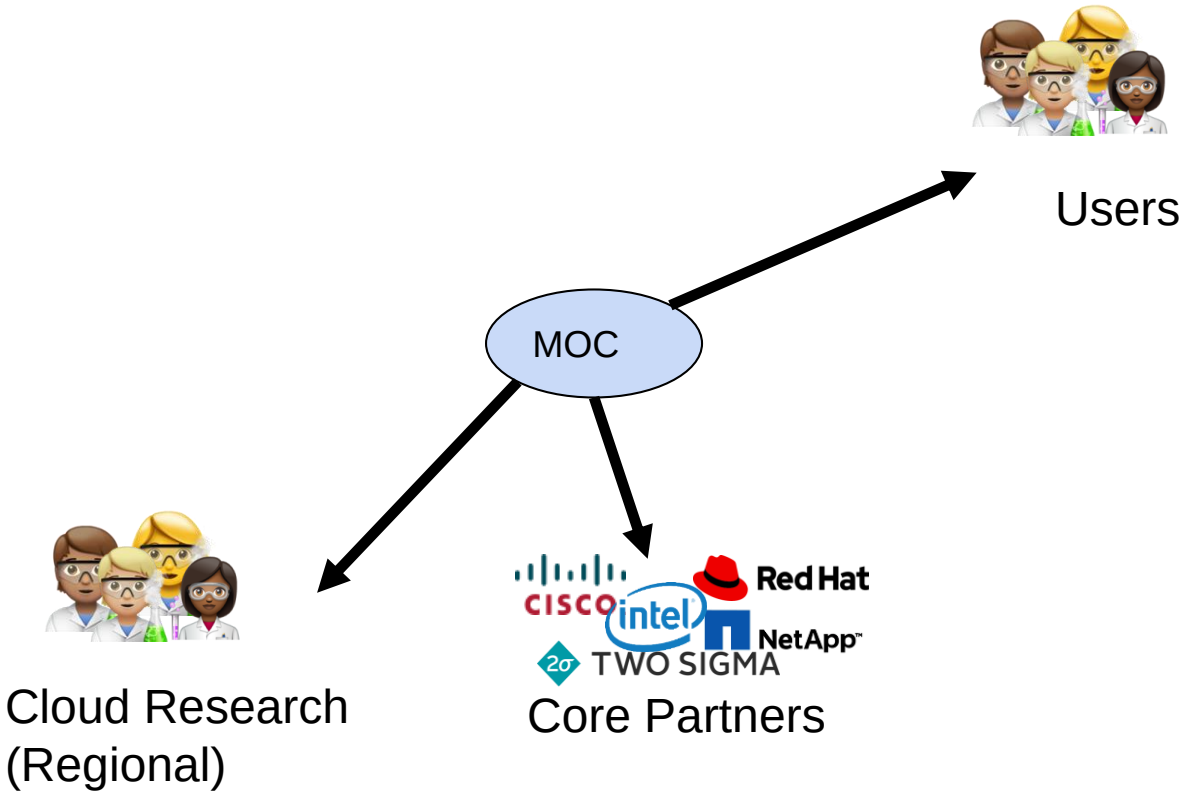
BOSTON UNIVERSITY

Cloud Computing

- We are creating, starting with the MGHPCC data center and Mass Open Cloud (MOC), a shared cloud that:
  - is more economical for research users than today's public clouds
  - enables cloud research
  - provides strong incentive for industry to engage, demonstrate innovation in a public venue, evaluate new technologies

- This has led to and taken advantage of large NSF/NIH/Industry grants:
  - MACS, NESE, Red Hat Collaboratory@BU, two hundred donated servers from Two Sigma...

Current MOC Model



1+ PB

2500 cores, ~40TB RAM

Elastic Secure Infrastructure

USAF

MIT LL

Red Hat

Two Sigma

Lenovo

Dell

Intel

Cisco

- Block and S3 Object storage
- Bare Metal Physical machines
- IaaS – VM, Volume
- Spark, Hadoop
- OpenShift: enterprise deployment of Kubernetes container platform:
- Built in CI, Monitoring, Load Balancing,

1+ PB

2500 cores, ~40TB RAM

Elastic Secure Infrastructure

USAF

MIT LL

Red Hat

Two Sigma

Lenovo

Dell

Intel

Cisco

OPEN DATA HUB
AI Platform powered by Open Source

ceph

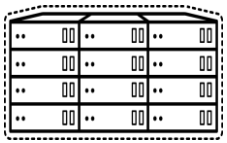
openstack.

RED HAT OPENSHIFT Container Platform

USAF
MIT LL
Red Hat
Two Sigma
Lenovo
Dell
Intel
Cisco
IBM

Elastic Secure Infrastructure

1+ PB



2500 cores, ~40TB RAM

400 Power9 Cores, **40** GPUs, **5**TB RAM

POWER9 AC922 servers

- 5.6x CPU to GPU BW vs standard Intel via NVLink 2.0
- 40 NVIDIA Tesla V100 GPUs delivering up to **5,000 Teraflops** for Deep Learning
- All major open source Deep Learning Frameworks
- PowerAI Distributed Deep Learning enables AI across multiple servers

HARVARD Dataverse

OPEN DATA HUB
AI Platform powered by Open Source

USAF
MIT LL
Red Hat
Two Sigma
Lenovo
Dell
Intel
Cisco
IBM
Harvard IQSS

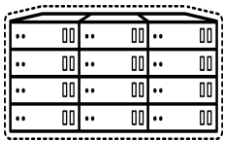
ceph

openstack.

RED HAT® OPENSHIFT Container Platform

Elastic Secure Infrastructure

1+ PB



2500 cores, ~40TB RAM

400 Power9 Cores, **40** GPUs, **5**TB RAM

- Current HDV hosted on AWS
  - 81,000 Datasets
  - 490,000 Files
  - 5.8 Million Downloads
- Moving to the MOC

HARVARD Dataverse

OPEN DATA HUB
AI Platform powered by Open Source

USAF
MIT LL
Red Hat
Two Sigma
Lenovo
Dell
Intel
Cisco
IBM
Harvard IQSS

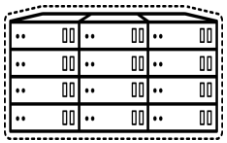
ceph

openstack.

RED HAT® OPENSHIFT Container Platform

Elastic Secure Infrastructure

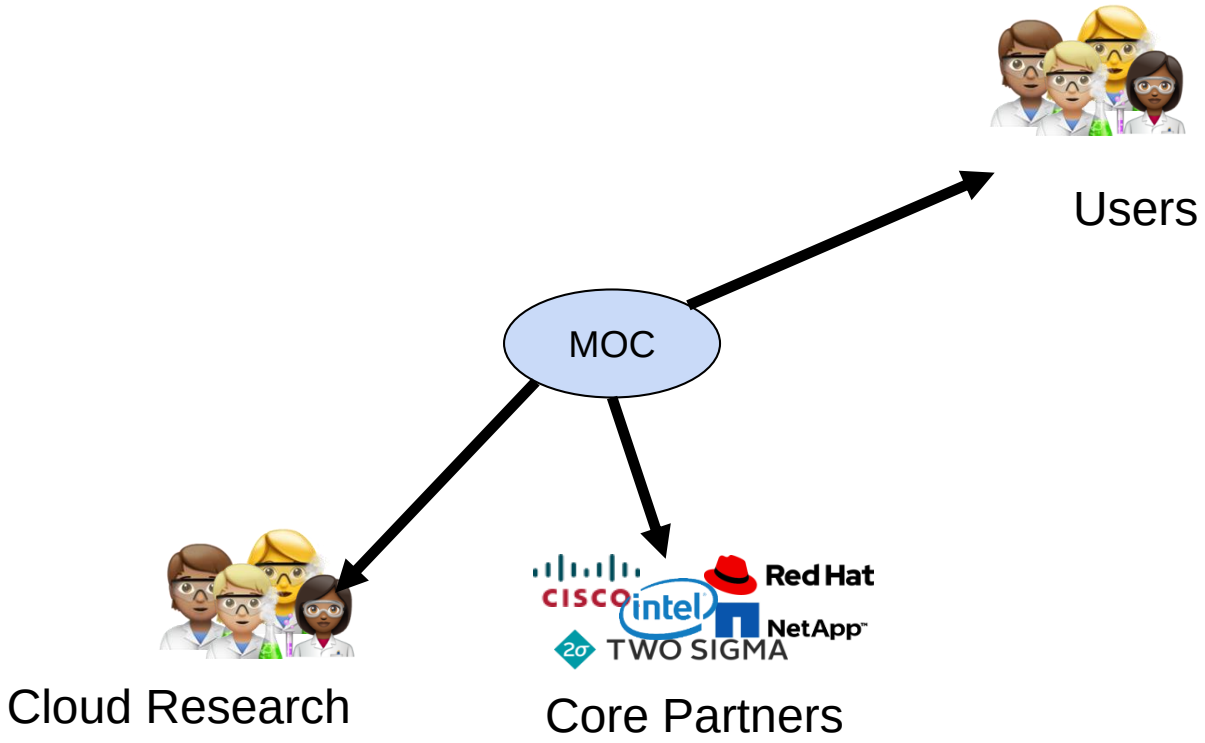
1+ PB

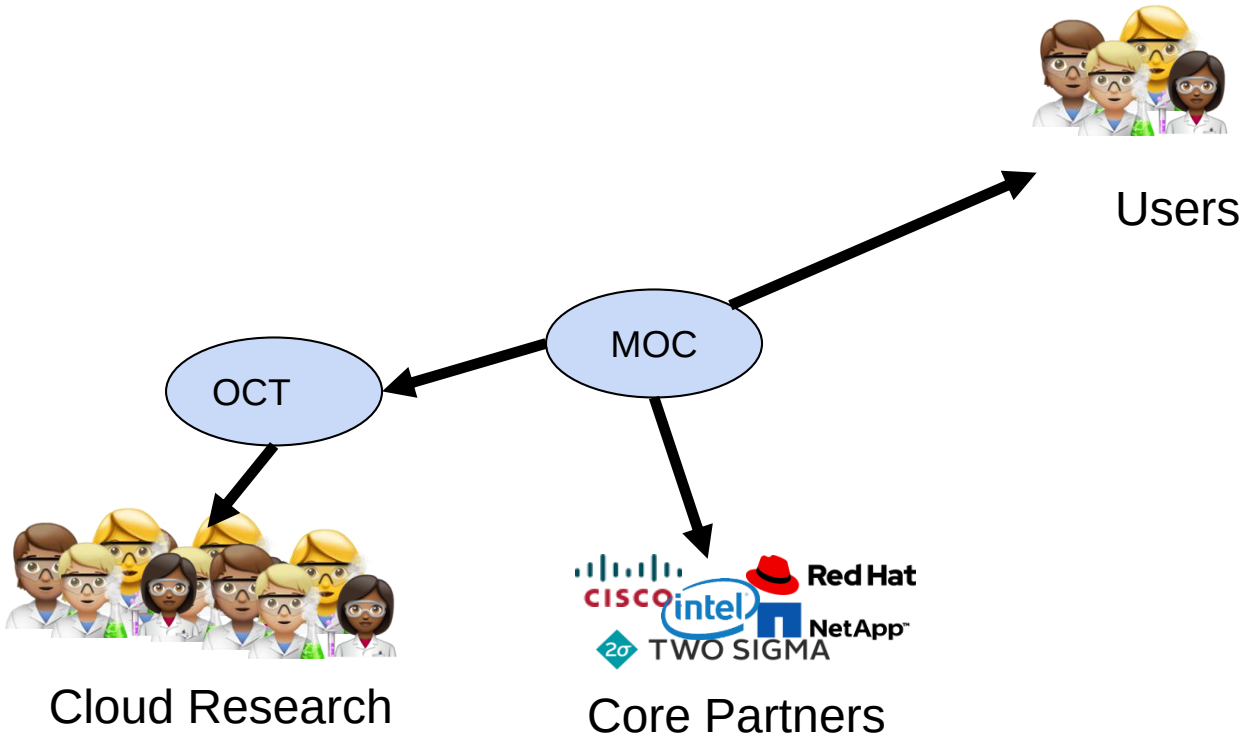


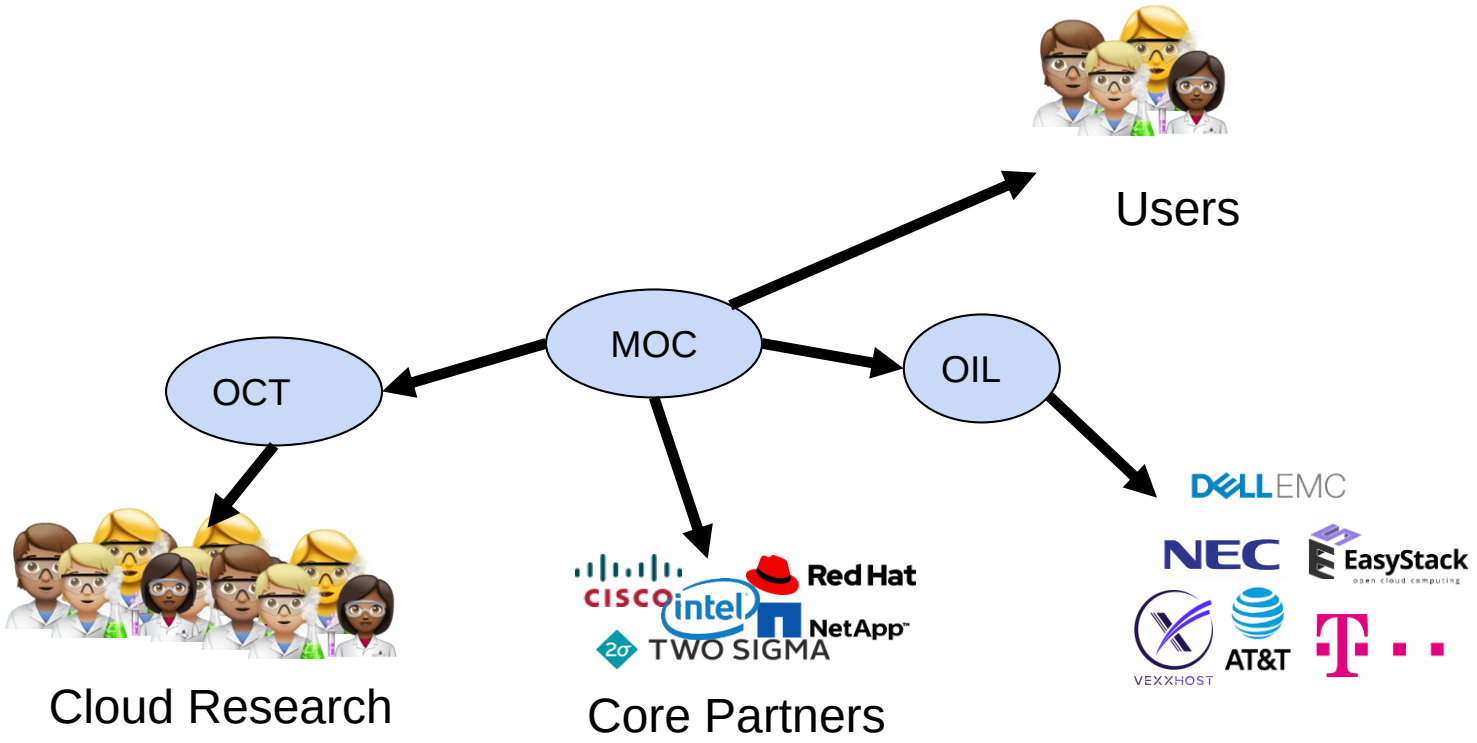
2500 cores, ~40TB RAM

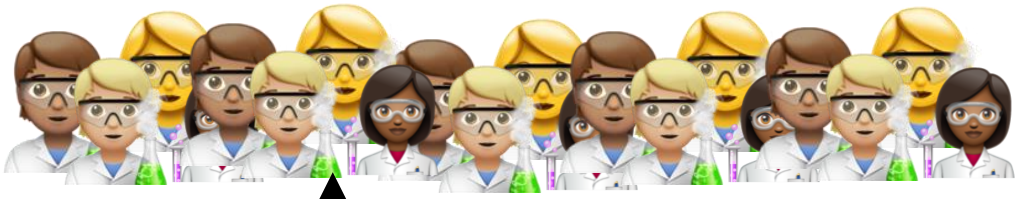
400 Power9 Cores, **40** GPUs, **5**TB RAM

- New North East Storage Exchange (NESE)
  - 20 PB + file system & Object storage
  - Massive data lake for region, co-located with MOC
  - Fraction of the cost of AWS S3


Users
MOC
Cloud Research
Core Partners
cisco
intel
Red Hat
NetApp
TWO SIGMA


Users
MOC
OCT
Cloud Research
Core Partners
cisco
intel
Red Hat
NetApp
TWO SIGMA


Users
MOC
OCT
OIL
Cloud Research
Core Partners
DELL EMC
NEC
EasyStack
AT&T
VEXXHOST



NERC

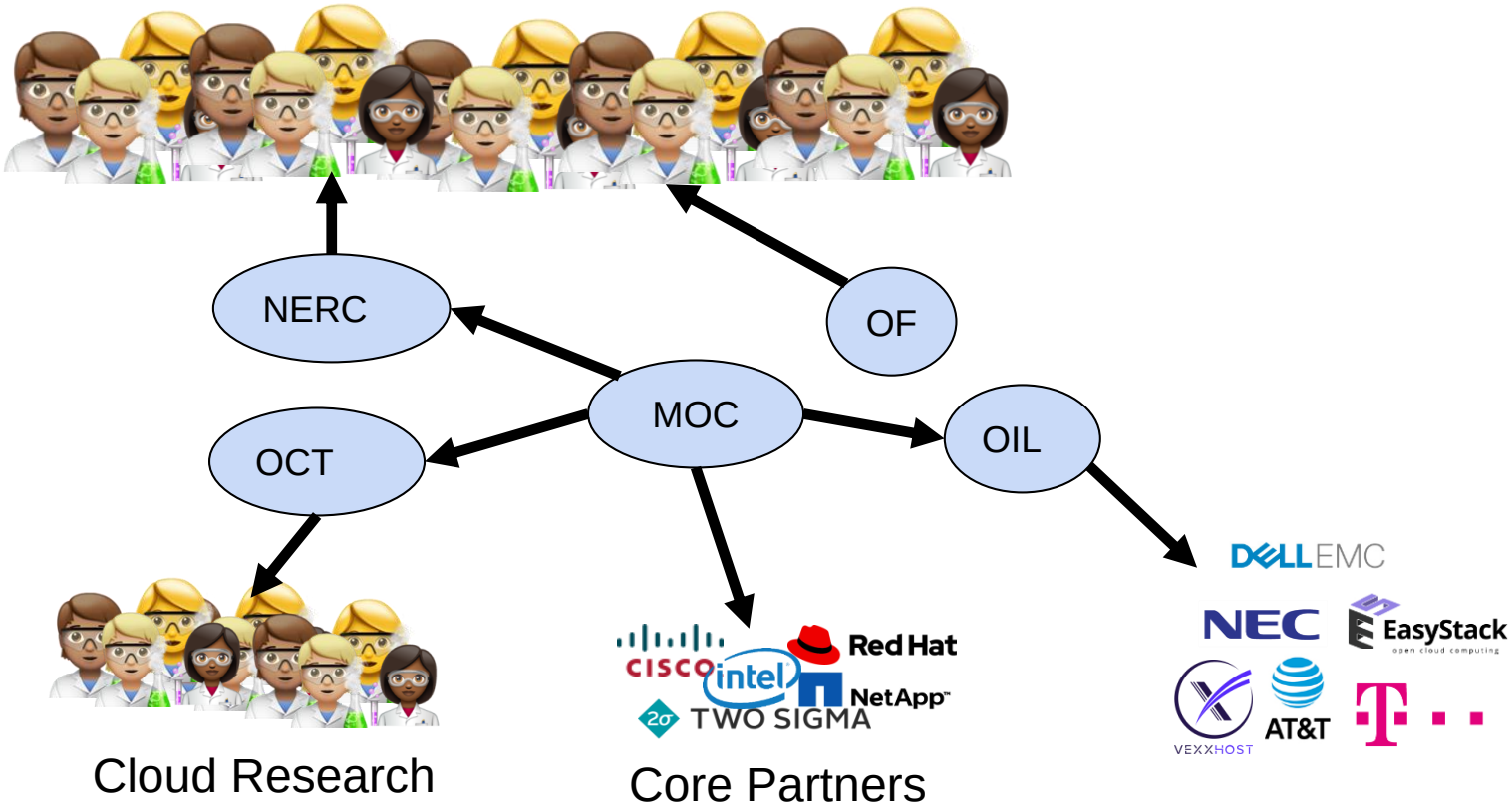
MOC

OCT

OIL

Cloud Research

Core Partners


NERC
OF
MOC
OCT
OIL
Cloud Research
Core Partners
DELL EMC
NEC
EasyStack
VEXXHOST
AT&T
Red Hat
NetApp
TWO SIGMA

Conclave Cloud Dataverse: Protected Computing in the Datacenter

Mayank Varia

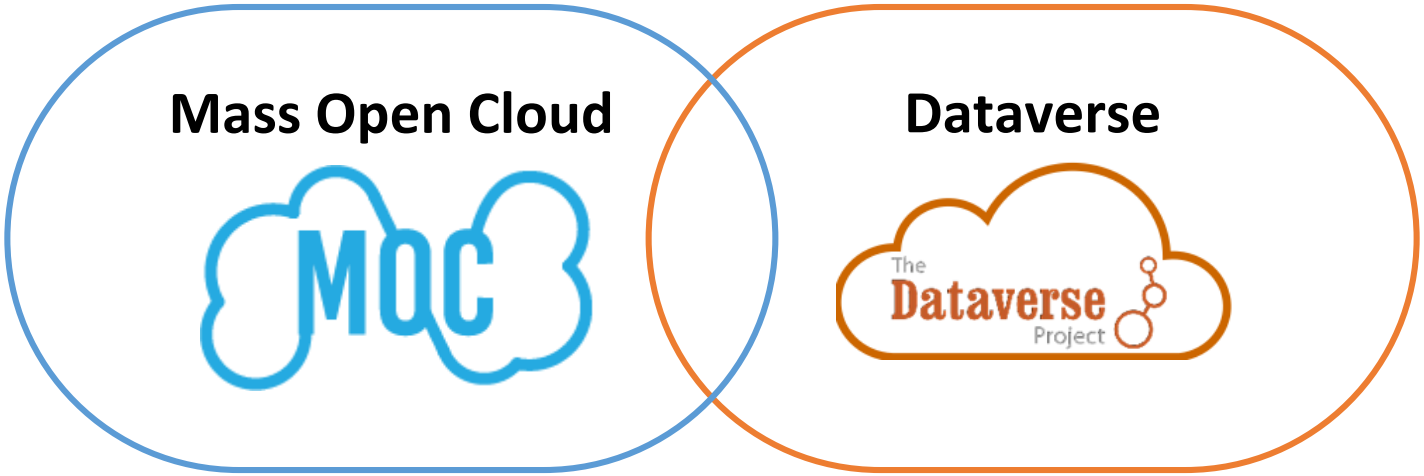
Boston University Hariri Institute for Computing

Privacy vs. utility on the cloud

- Companies in MA want to *compute average salary differences* by gender and race, *without exposing average salary of any company*
- Tier-1 trauma centers in Boston want to *generate aggregate reports* about cases they service *without revealing any patient data*
- Researchers in hospitals want to generate aggregate statistics about rare diseases across multiple hospitals *without revealing patient data*

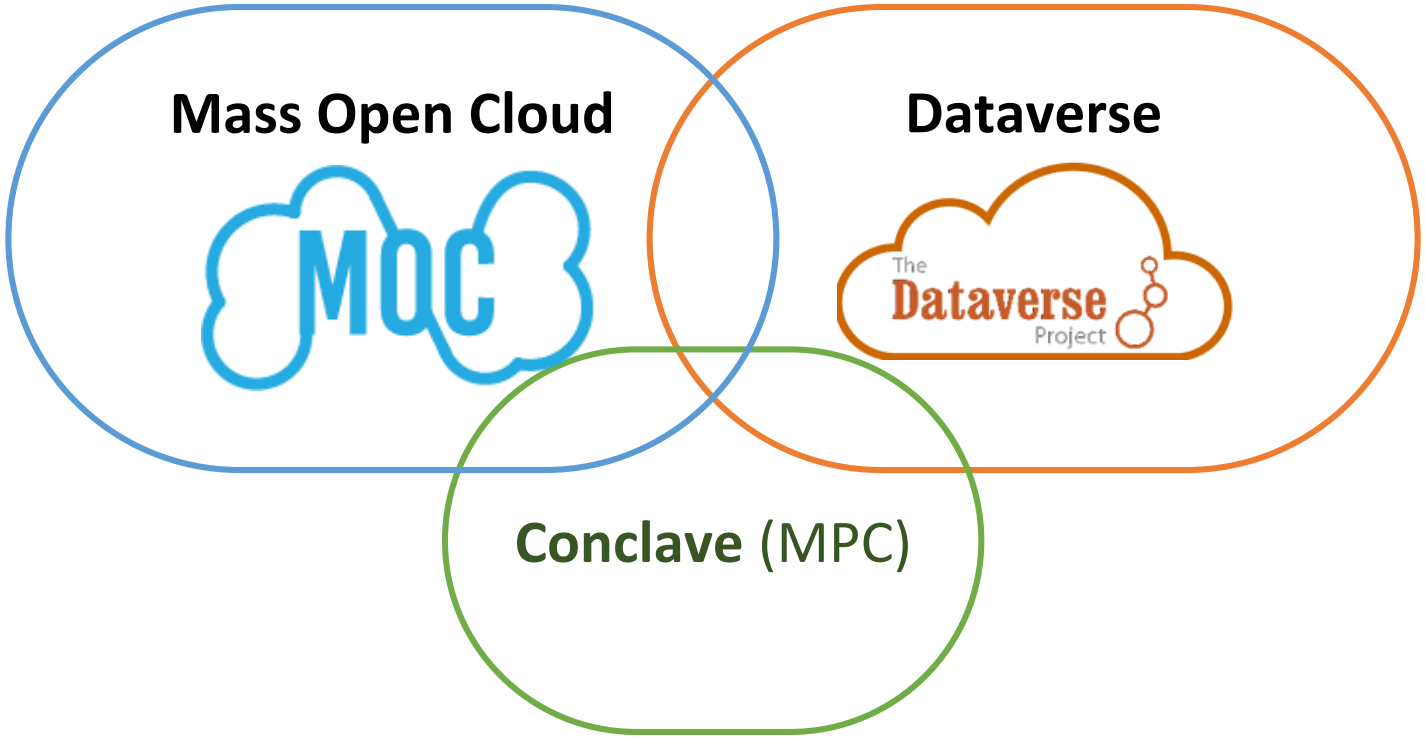
Common theme: organizations want to run data analytics in the public cloud, but do not trust a single public cloud provider

Privacy-preserving **scientific analysis in an open cloud**



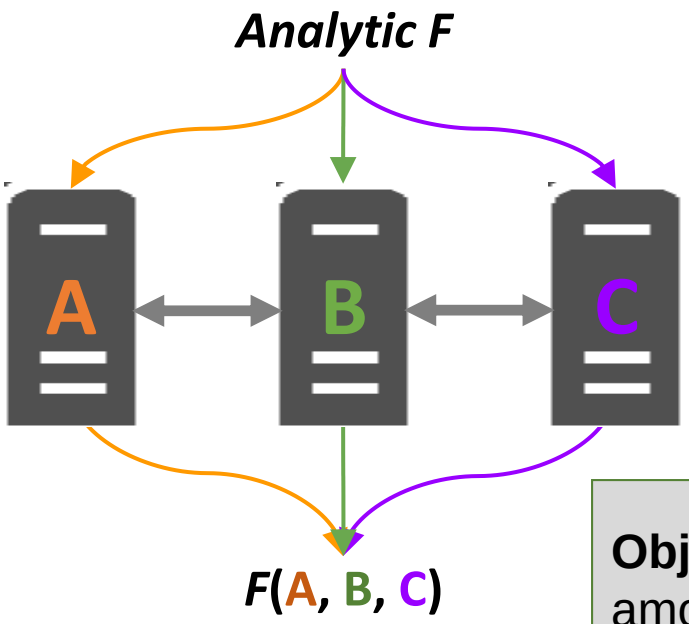
Cloud Dataverse combines the power of cloud computing and storage with access to thousands of datasets from a feature-rich repository platform

Privacy-preserving scientific analysis in an open cloud

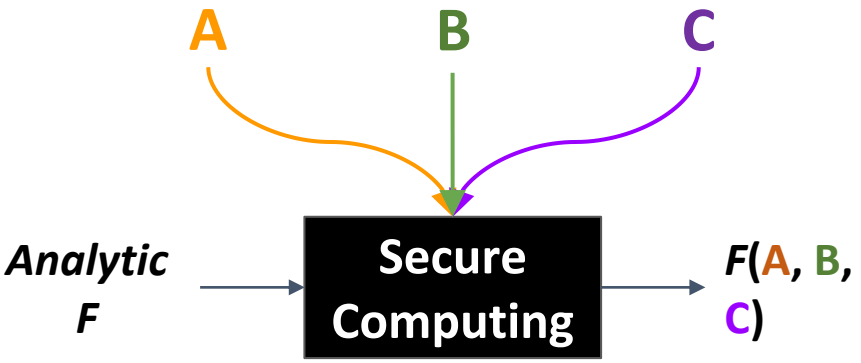


Secure multi-party computation (MPC)

Makes this…



…look as if it were this

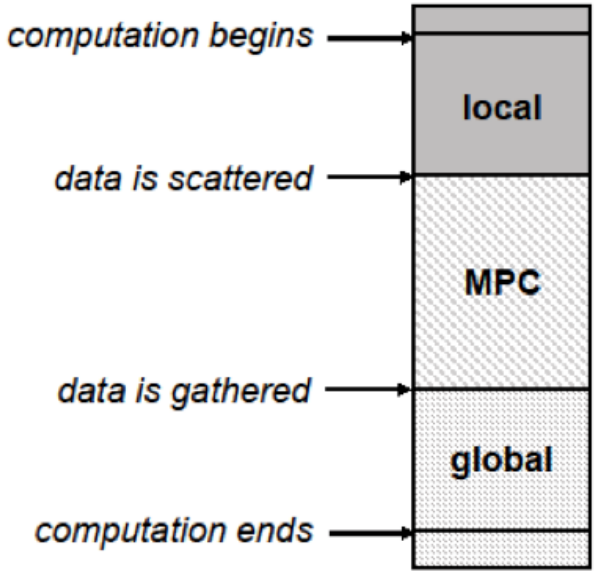


Objective: Federate trust in data & computing among several compute entities

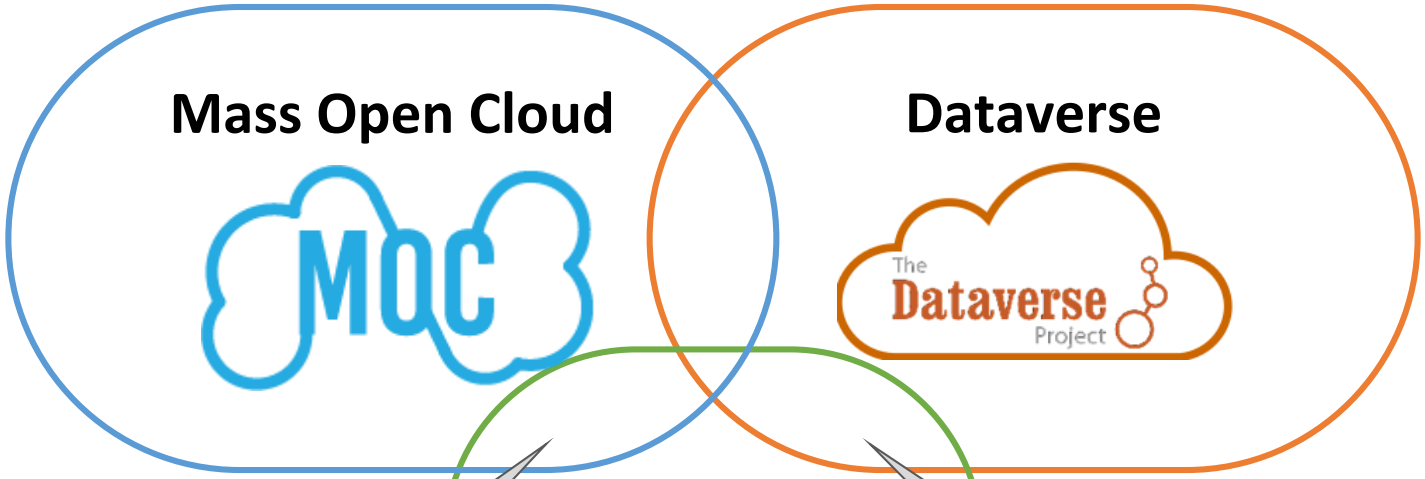
Tradeoff: Gain security at expense of networking

Conclave: MPC at scale

- SQL-like programming language
  ⇒ *No MPC experience needed*
- Static analysis to discern boundaries of secure computing
  ⇒ *Scale to ~billions of records*
- Dispatcher executes jobs on available backends
  ⇒ *No new infrastructure*



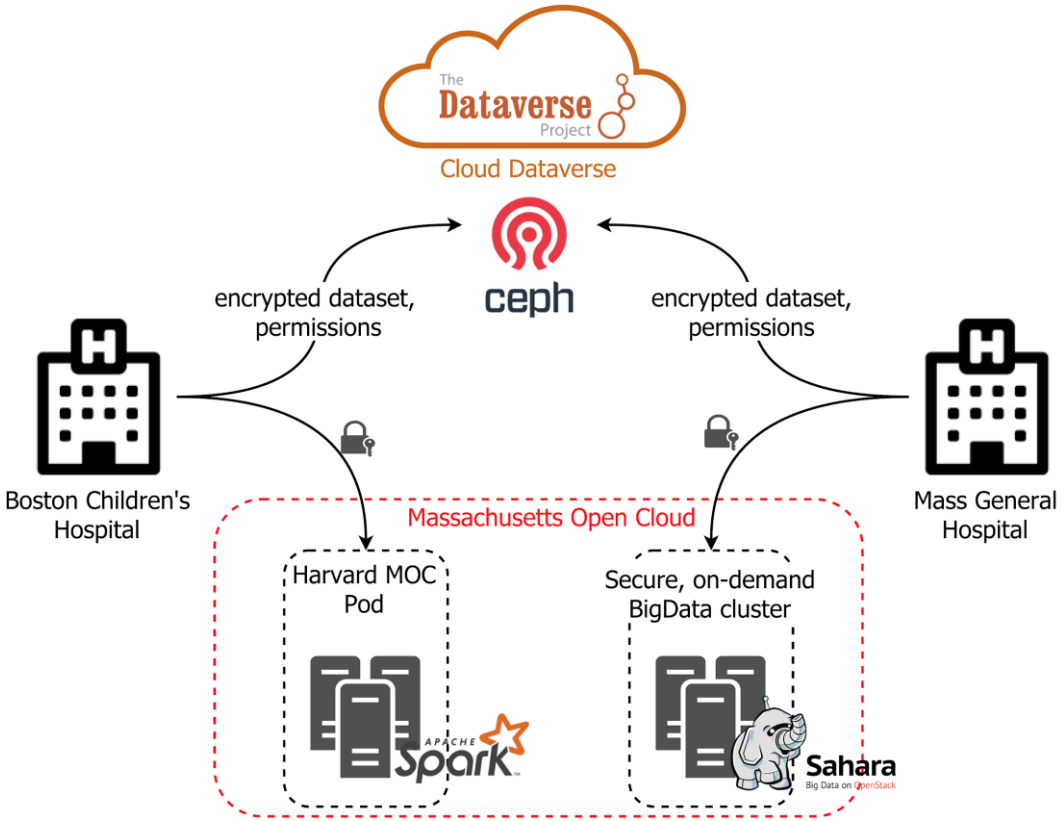
The synergistic payoff



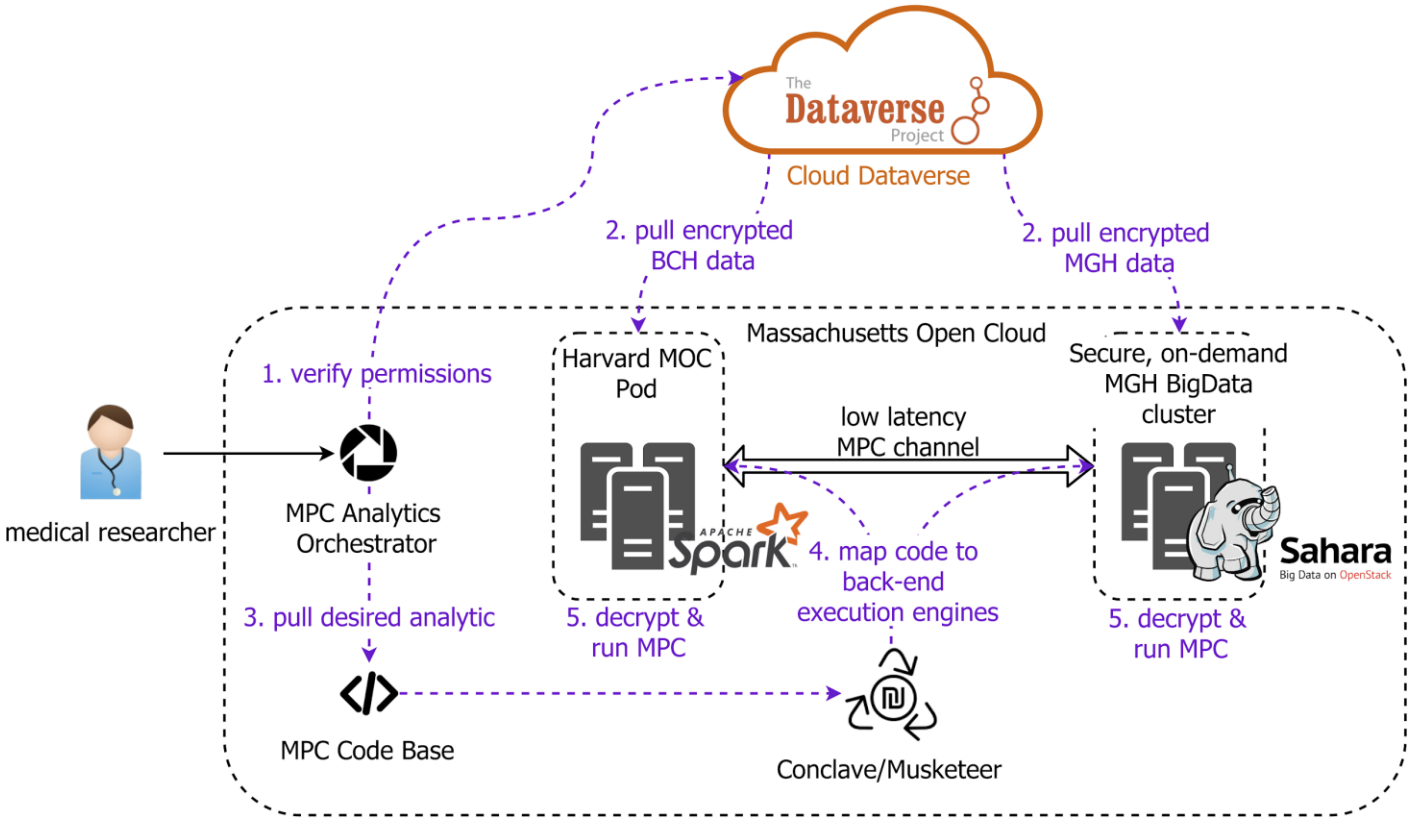
Benefit: Bring protected computing to existing datasets

Benefit: Leverage open cloud to improve performance

Conceptual workflow: data upload



Conceptual workflow: data analysis



MPC takeaway: we can have it both ways

We can derive knowledge (K) from data held by several organizations, without sharing it or trusting any third party

$$K = f(\text{  }, \text{  }, \text{  })$$

Thanks!

github.com/ccd-mpc
github.com/multiparty/conclave

AI PRESENTATION

Kate Saenko

Boston University Hariri Institute for Computing

BOSTON UNIVERSITY

Artificial Intelligence

Large group of faculty and students at BU working on AI, in particular:

- Deep learning
- Supervised, unsupervised, semi-supervised learning
- Active learning
- Domain adaptation
- Embeddings
- Encoder/decoder models
- Adversarial networks

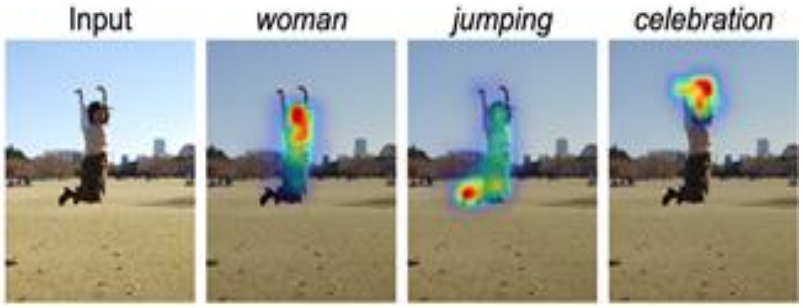
Artificial Intelligence for

Image Credit: Shutterstock/Sepp photography

- Climate change & digital media
- Animal behavior analysis
- Image and video retrieval & semantic hashing
- Human gesture and activity recognition in video
- Smart crowdsourcing of image annotations & humans-in-the-loop systems
- Visual and language understanding, e.g., teaching machines to “talk” about what they see, recognizing text in images
- Home-based physical therapy, assistive systems for users with disabilities
- Detecting and remediating racial and gender bias
- COVID prediction

Artificial Intelligence: Making learning methods for computer vision accountable & interpretable

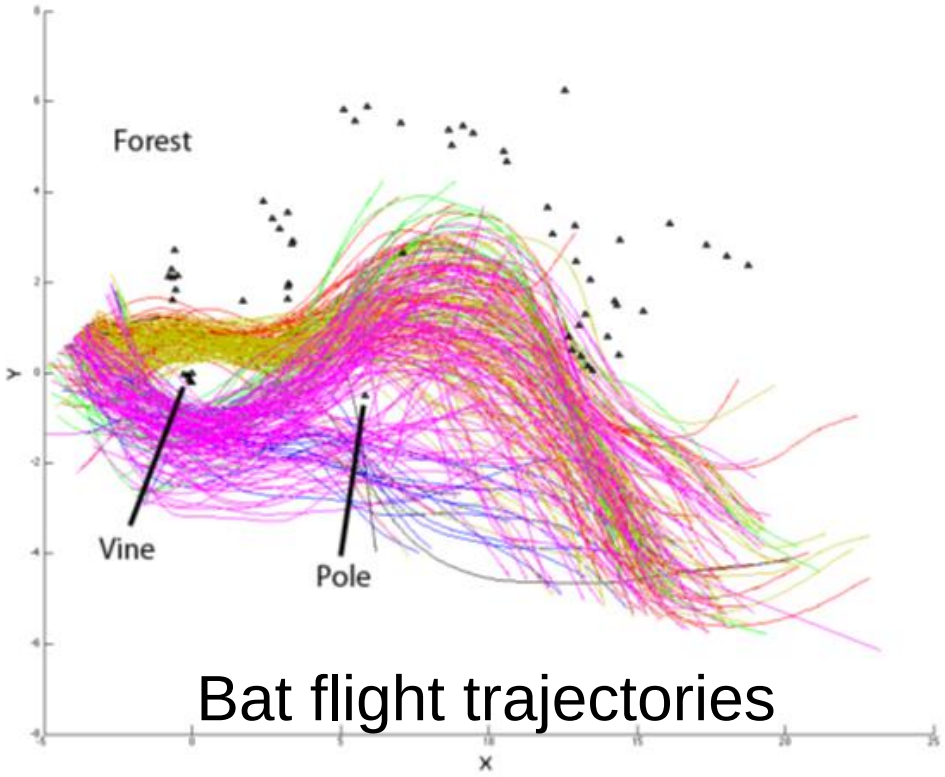
- Can machine learning methods explain their outputs?
- Can we visualize what evidence and what parts of the model support a particular conclusion?
- Can an algorithm explain what it “sees”?



BOSTON UNIVERSITY

Artificial Intelligence for Animal Behavior Analysis

Watch the video of 3D bat flight:
https://youtu.be/ntGJhyBhtdk



Bat flight trajectories

Artificial Intelligence Research

- Deep learning, artificial neural networks
- Computer vision, natural language processing
- Dataset bias, transfer learning

Understand images and language

A baseball game in progress with the batter up to plate

A man is riding a bicycle

Q: What is the child standing on?

A: skateboard

Find region for "window upper right"

Find when "girl looks up at the camera and smiles"

Multi-lingual image-text retrieval

Chinese: 一位男冲浪者在一个小波..

English: A surfer rides a small wave and has ..

Japanese: きれいな海でサー...

German: Ein Surfer reitet eine kleine Welle...

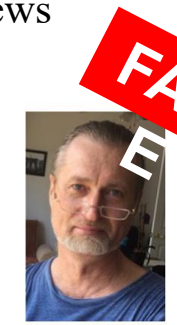
Detecting Neural Fake News

Man Who Jumped From Ambulance Says It's New York City's Fault.

Jon Vernick, the man who plunged from an ambulance parked on Manhattan's Upper East Side on Wednesday, says it was the city's fault for allowing him to get that close to the patient. Writing in the New York Post, Vernick — who miraculously survived the fall and was listed in stable condition with a broken collarbone — said that he was waiting for the doctor to arrive when he jumped out of the ambulance.

.
.
.

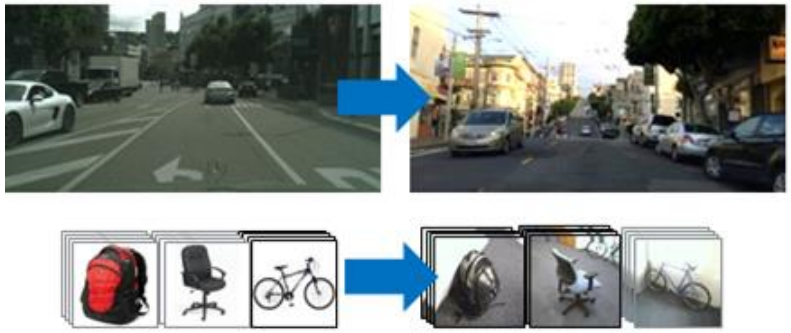
The EMS workers did not have the ability to do anything to stop me." A spokesman for City Council Health Committee chair Ydanis Rodriguez said the committee was in the process of conducting an investigation. Vernick says he went for an interview with NBC New York and tried to tell his story so the city would not continue to put people in precarious positions. Read



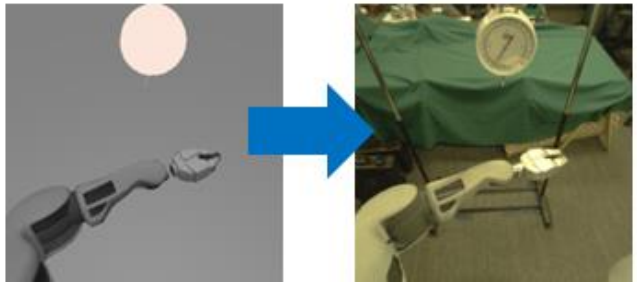
On June 11, 2016, Yaugeni Kralkin, who was drunk, exited an ambulance en route to Staten Island University Hospital South Campus.

Transfer knowledge and overcome dataset bias

From dataset to dataset



From simulated to real control

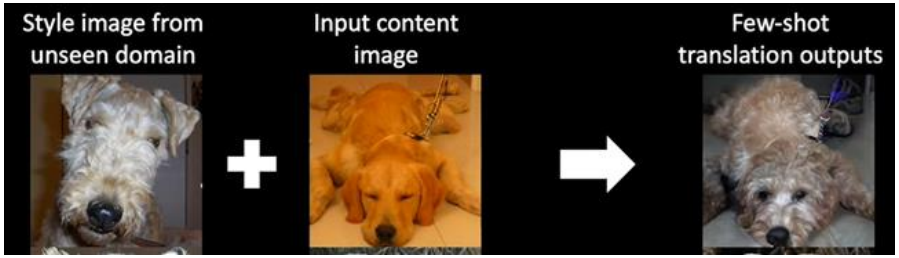


From web to robot

From one demographic to another

Deep reinforcement learning: sim2real

Generate realistic images, changing their style or content



Panel Discussion and Q&A

Get Involved

Leveraging the Computational Perspective in a Data-Driven World for a Better Society

Website: bu.edu/hic

Twitter: @BU_Computing

Facebook: BUcomputing