



**MIDAS**

MICHIGAN INSTITUTE  
FOR DATA SCIENCE  
UNIVERSITY OF MICHIGAN

# **The 2017 MIDAS Symposium**

## **A Data-Driven World: Potentials and Pitfalls**

**October 11  
Rackham Building &  
Michigan League**

[midas.umich.edu](http://midas.umich.edu)

Wireless: MGuest



**@ARC\_UM  
#umbigdata  
#umtalks**

THE VALUE OF PERFORMANCE.

***NORTHROP GRUMMAN***



## INFLUENCE

THE FUTURE OF COMPUTER  
INFORMATION SYSTEMS

At Northrop Grumman, our work with cutting-edge technology is driven by something human: the lives our technology helps protect.

We deliver critical information when and where it's needed to keep our military, intelligence, federal, state and local, international and commercial customers better informed, connected and secure.

Join us as we continue this important mission of providing meaningful solutions to our global customers.

Let's do the work that matters - together.

**CAREERS THAT TAKE YOU PLACES.**



[careers.northropgrumman.com](http://careers.northropgrumman.com)

@2016 Northrop Grumman is committed to hiring and retaining a diverse workforce. We are proud to be an Equal Opportunity/Affirmative Action Employer, making decisions without regard to race, color, religion, creed, sex, sexual orientation, gender identity, marital status, national origin, age, veteran status, disability, or any other protected class. U.S. Citizenship is required for most positions. For our complete EEO/AA and Pay Transparency statement, please visit [www.northropgrumman.com/EEO](http://www.northropgrumman.com/EEO)

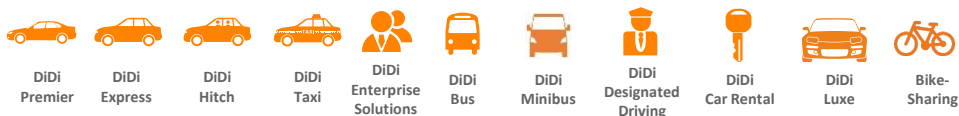


**DiDi Chuxing** is the world's leading mobile transportation platform. We are committed to working with communities and partners to solve the world's transportation, environmental and employment challenges by using big data-driven deep-learning algorithms that optimize resource allocation. By continuously improving user experience and creating social value, we strive to build an open, efficient, collaborative, and sustainable transportation ecosystem.

**2022 MISSION:** To become a global leader in smart transportation and automotive technology, the world's largest operator of vehicle networks and a global leader in smart transportation systems.

## SERVICES

DiDi Chuxing offers a diverse range of transportation services through one mobile app:



## TECHNOLOGICAL ADVANTAGES

**70TB+** new data/day

**4,500TB+** data processed/day

**20bn+** routing requests/day

**14bn** location points/day

- Intelligent ride-matching system enables multi-person carpooling within and across cities
- Real-time traffic mapping and route optimization
- DiDi supports cities in efficient and sustainable transportation planning with big data-based transportation capacity

## OUR GROWTH

**+400 million**  
passengers

**>20 million**  
rides completed per day

**> 17.50 million**  
job opportunities provided

**~7,000**  
employees (~ 50% engineers & data scientists)



## OUR VERTICALS

**Taxi:** 1.68 million drivers; partnering with over 200 taxi companies

**Private Car:** DiDi Premier and DiDi Express cover 400+ cities with an ExpressPool option

**Hitch (Social Ridesharing):** 2.2 million daily rides at the peak

**Designated Driving:** in over 200 cities

**Enterprise Solutions:** 60,000+ corporate clients

**Car Rental:** in over 1,500 cities across over 100 countries

**Uber China:** in over 400 cities

## MILESTONES

DiDi Dache and Kuaidi Dache were founded in 2012 as taxi-hailing apps and later developed private-car hailing business.

DiDi Dache and Kuaidi Dache merged in February 2015, later rebranded as DiDi Chuxing. In August 2016, DiDi acquired Uber China.

As of January 2017, DiDi has built investment and technology partnerships with five leading rideshare companies of the world, including Lyft, Grab, Ola, Uber, and 99.

As the world's leading mobile transportation platform, DiDi is well capitalized by world-class strategic and financial investors, including Tencent, Alibaba and Apple.

# Symposium Schedule

**7:30 a.m.**

## **Check-in and Coffee**

**8:30 a.m.**

## **Welcome**

Eric Michielssen, Associate Vice President, Research Computing  
Brian Athey & Al Hero, MIDAS Co-Directors

**8:45 a.m.**

## **Statistical Methods for Problems in Biology**

Daniela Witten, Associate Professor of Statistics and Biostatistics  
University of Washington

**9:45 a.m.**

## **Analyzing Words to Understand People**

James Pennebaker, Professor of Psychology  
University of Texas at Austin

**10:45 a.m.**

## **MIDAS Challenge Research Initiatives Panel**

Presentations from researchers on MIDAS-supported projects in transportation, learning analytics, health science and social science, including time for questions and answers.

Panelists: Anna Gilbert, Mathematics | Carol Flannagan, U-M Transportation Research Institute | Rada Mihalcea, Electrical Engineering and Computer Science | Trivellore Raghunathan, Survey Methodology | Srijan Sen, Molecular and Behavioral Neuroscience Institute, Psychiatry

**Noon**

## **Lunch and Poster Competition**

Located in the Michigan League

See map on page 10 for directions to the Michigan League, and details on the poster competition

**1:30 p.m.**

**MIDAS/UMOR/Dissonance Keynote:**

**Weapons of Math Destruction**

Cathy O'Neil, data scientist and author

**2:30 p.m.**

**Can Data Science Save the Environment?**

Francesca Dominici, Professor of Biostatistics

Harvard T.H. Chan School of Public Health

**3:30 p.m.**

**Computer Science and \_\_\_\_\_: Better Together, The Value of An Interdisciplinary Approach**

Nadya Bliss, Director, Global Security Initiative

Arizona State University

**4:15 p.m.**

**Question & Answer Session**

Daniela Witten | James Pennebaker | Cathy O'Neil | Nadya Bliss

**4:45 p.m.**

**Challenges and Opportunities for Industry**

Henry Liu, Didi Chuxing

Sheikh Shuvo, Mighty AI

Brian Letort, Northrop Grumman

Beaumont Vance, TD Ameritrade

**5:30 p.m.**

**Closing Remarks and Student Poster Awards**

Brian Athey and Al Hero, MIDAS Co-Directors

Eric Schwartz (Ross School of Business)

Ivo Dinov (School of Nursing)

**6 p.m.**

**Reception and Open House**

MIDAS Offices, Weiser Hall, 500 Church St., Sixth Floor

See map on page 9 for directions to Weiser Hall

## Weapons of Math Destruction

Cathy O’Neil, data scientist and author



**Abstract:** We live in the age of the algorithm. Increasingly, the decisions that affect our lives—where we go to school, whether we get a car loan, how much we pay for health insurance—are being made not by humans, but by mathematical models. In theory, this should lead to greater fairness: Everyone is judged according to the same rules, and bias is eliminated.

But as Cathy O’Neil reveals, the opposite is true. The models being used today are opaque, unregulated, and uncontestable, even when they’re wrong. Most troubling, they reinforce discrimination: If a poor student can’t get a loan because a lending model deems him too risky (by virtue of his zip code), he’s then cut off from the kind of education that could pull him out of poverty, and a vicious spiral ensues. Models are propping up the lucky and punishing the downtrodden, creating a “toxic cocktail for democracy.” Welcome to the dark side of Big Data.

Tracing the arc of a person’s life, O’Neil exposes the black box models that shape our future, both as individuals and as a society. These “weapons of math destruction” score teachers and students, sort résumés, grant (or deny) loans, evaluate workers, target voters, set parole, and monitor our health. O’Neil calls on modelers to take more responsibility for their algorithms and on policy makers to regulate their use. But in the end, it’s up to us to become more savvy about the models that govern our lives.

**Bio:** Cathy O’Neil is a data scientist and author of the blog [mathbabe.org](http://mathbabe.org). She earned a Ph.D. in mathematics from Harvard and taught at Barnard College before moving to the private sector, where she worked for the hedge fund D. E. Shaw. She then worked as a data scientist at various start-ups, building models that predict people’s purchases and clicks. O’Neil started the Lede Program in Data Journalism at Columbia and is the author of *Doing Data Science*. She is currently a columnist for Bloomberg View.

Courtesy of:



## Featured Speakers

### **Daniela Witten, Associate Professor of Statistics and Biostatistics, University of Washington**

*Statistical Methods for Problems in Biology*

As the pace and scale of data collection increase across all areas of biology, there is a growing need for effective and principled statistical methods for data analysis. Two projects in my lab help fill this gap. First, calcium imaging is transforming the field of neuroscience by making it possible to assay the activities of large numbers of neurons simultaneously. The “fluorescence trace” of each neuron needs to be de-convolved into spike times. I will discuss the optimization approach in an auto-regressive model for the deconvolution. Second, it is increasingly common to collect “multi-view data”: multiple types of data (e.g. DNA sequence, clinical measurements) on a single set of observations (e.g. patients). We will consider the clustering problem in such datasets.



### **James Pennebaker, Professor of Psychology, University of Texas at Austin**

*Analyzing Words to Understand People*

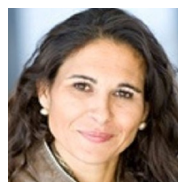
Function words include pronouns, prepositions, articles, and other common but almost-invisible words in most languages. Their use often signals speakers’ relationships with their audience, their subject matter, and how the speakers think and feel about themselves. Multiple studies find function words can reveal personality, emotional state, deception, status, and thinking styles. Implications for research in medicine, marketing, law, education, literature, and other disciplines are discussed. Oh yes, language use is also relevant to recent and historical political trends.



### **Francesca Dominici, Professor of Biostatistics, Harvard T.H. Chan School of Public Health**

*Can Data Science Save the Environment?*

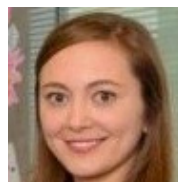
How would you react if I give you bulletproof evidence that a jumbo jet will be hijacked and crashed every 12 days? The threat of losing human lives at this scale is actually real, except that it comes from climate change and air pollution, instead of terrorists. We have used an artificial neural network model with on-the-ground and satellite-based data to estimate pollution levels across the continental U.S. We have paired such data with health record data and developed statistical methods and algorithms for analysis. We have shown that air pollution is killing 12,000 senior citizens each year—the equivalent of a jumbo jet crashing every 12 days, and that federal limits on the nation’s most widespread air pollutants are not stringent enough.



### **Nadya Bliss, Director, Global Security Initiative, Arizona State University**

*Computer Science and \_\_\_\_\_: Better Together; The Value of an Interdisciplinary Approach*

Internet connectivity has fundamentally changed our lives. The data we collect and analyze presents new opportunities and the potential to transform application domains. However, it is also clear that the persistence of connectivity has created a number of challenges. My claim is that many of those challenges can be attributed to the fact that technological advancement has often occurred in silos. Stronger collaboration between computer science and other disciplines will help mitigate these challenges. We have integrated the principles of interdisciplinarity at ASU’s Global Security Initiative – both from perspectives of organizational design and research approaches.



## WHY TD AMERITRADE IS A GREAT PLACE TO WORK

What does it mean to be *client-centric* and put *integrity first*? It means focusing on people rather than just the bottom line. It means investing in our communities, as well as technology and employee development. And ensuring our values resonate in the relationships we build and the work we do. To that end, we actively foster a work environment that empowers everyone to do their best and drives continuous improvement, problem solving and reduces complexity.

TD Ameritrade (NASDAQ: AMTD) offers investing and trading services for more than seven million funded client accounts that total more than \$800 billion in assets and custodial services for more than 5,500 independent registered investment advisors (RIAs). From first-time investors to sophisticated traders, TD Ameritrade is focused on helping individuals seek out their financial goals.

As a leading employer in financial services, we believe in building strong relationships, achieving quality results and working with integrity. It's at the heart of everything we do. That's why we provide you with comprehensive training, benefits and programs that inspire you to achieve more every day.

### Join our team and experience:

- Outstanding culture where *people matter* and diverse backgrounds and perspectives are encouraged
- Job-specific business and technical training
- Job mobility and growth opportunities
- Competitive Total Rewards including quarterly bonus incentives, 401k plan with company matching, profit sharing, paid time off, medical, vision and dental insurance and back up child care
- Commitment to ongoing innovation, client education, and maintaining outstanding service levels
- Tuition reimbursement program

If you are interested in learning more about us, visit our careers site <http://careers.tdameritrade.com/>.

For further questions please contact us at  
866-373-3872 or  
[ZHRSC-AssociateSupport@tdameritrade.com](mailto:ZHRSC-AssociateSupport@tdameritrade.com).



TD Ameritrade Holding Corporation (NASDAQ: AMTD). TD Ameritrade is a trademark jointly owned by TD Ameritrade IP Company, Inc. and The Toronto-Dominion Bank. © 2017 TD Ameritrade IP Company, Inc. This communication is not an offer or solicitation for brokerage services or investment advisory services from TD Ameritrade, Inc., member FINRA/SIPC, a wholly owned subsidiary of TD Ameritrade Holding Corporation.

TD Ameritrade is an Equal Opportunity Employer.

Distributed by: TD Ameritrade Holding Corporation, 200 South 108th Avenue, Omaha, NE 68154-2631.



# Mercury Fund

**Mercury Fund** is an early-stage venture capital firm. With over \$200 million under management, Mercury focuses on entrepreneurs and technology innovation originating in the U.S. Midcontinent. Our investment themes target SaaS, Cloud, and Data Science/AI platforms that make the industrial ecosystems of Middle America more competitive and efficient.

Over a decade ago, Mercury recognized the rise of entrepreneurs and innovation in the underinvested Midcontinent. Our venture investment strategy leverages our unparalleled network of midcontinent startup development organizations, corporate innovation partners, and co-investors to assist entrepreneurs with the resources they need to rapidly scale their businesses. We have spent our professional careers advising, mentoring, and investing in Mid-American entrepreneurs. We believe now, more than ever, the Midcontinent represents a great venture capital opportunity.

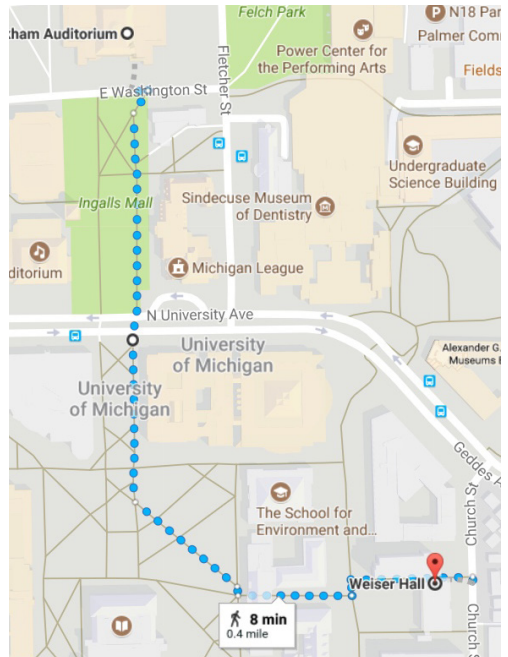
[mercurymfund.com](http://mercurymfund.com)

## Map to MIDAS Offices Weiser hall, 500 Church St.

For open house and reception  
at 6 p.m.



Scan for Google Map



# MIDAS Student Poster Competition

Posters from more than 60 graduate students and faculty members are featured in the MIDAS Poster Competition (in the Michigan League).

Student poster award categories:

- Most Innovative Use of Data
- Most Likely Societal Impact
- Most Interesting Methodological Advances
- Most Likely Transformative Scientific Impact
- Most Likely Health Impact
- Best Overall

Research areas represented:

- Health Science
- Learning Analytics
- Transportation
- Urban Planning
- Security
- Marketing
- Geophysics
- Energy
- Social Science

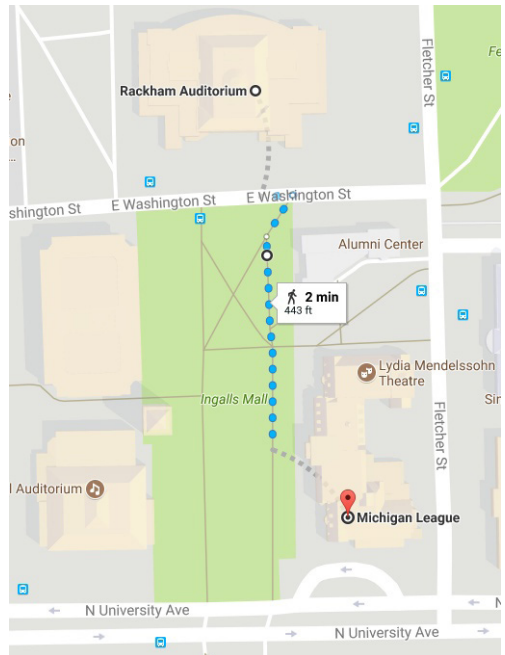
---

## Map to Michigan League 911 N. University Ave.

For lunch and poster session  
at noon



Scan for Google Map



# SAVE THE DATE

## 2017 U-M Data Science Research Forum

8:30 a.m. - 5:30 p.m. | December 1 | Michigan League

You're invited to the 2017 U-M Data Science Research Forum, which will feature faculty and student presentations on data science research across the University.

### Forum highlights

- Oral and poster presentations on
  - Theoretical foundations of data science
  - Data science methodology
  - Data science applications in any research domain
  - Social impact of data science research
- Industry engagement workshop
  - Adrian Fortino, Partner, Mercury Fund
  - Mike Psarouthakis, Director, U-M Venture Center
  - Kevyn Collins-Thompson, Associate Professor, U-M School of Information
  - Mike Cafarella, Associate Professor, U-M Computer Science and Engineering
- Presentations on data science infrastructure and consulting services
  - Brock Palen, Director of Advanced Research Computing - Technology Services
  - Kerby Shedden, Director of Consulting for Statistics, Computing and Analytics Research
- Keynote by Chris Rozell, Associate Professor, Electrical Engineering and Computer Engineering, Georgia Institute of Technology, "Closing the Loop Between Mind and Machine: Building Algorithms to Interface with Brains at Multiple Stages"
- Networking Reception

All presentations will come from submissions from the U-M data science research community (faculty, staff, trainees).

**Abstract Submission Deadline: October 23, 2017.**

For details and to submit an abstract, see [midas.umich.edu/forum](http://midas.umich.edu/forum).



# MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN

The Michigan Institute for Data Science (MIDAS) creates and maintains a thriving community in data science on the U-M campus in the service of its University of Michigan stakeholders and of emerging national needs.

MIDAS oversees and leads Data Science Challenges in four initial areas:

- Transportation Research
- Learning Analytics
- Personalized Medicine and Health
- Social Science

MIDAS brings data science faculty together to foster a collaborative, cross-cutting methodological and applied research community in data science. MIDAS leads and coordinates next-generation data science education and training; engages industry in data science at U-M; supports student groups including the Michigan Data Science Team (MDST), the Michigan Student Artificial Intelligence Lab (MSAIL), and the Computational Social Science Workshop (CSSW); and promotes usage of data science services and infrastructure on the U-M campus.

MIDAS is co-directed by Brian Athey (bleu@umich.edu) and Al Hero (hero@umich.edu). For general inquiries: [midas-contact@umich.edu](mailto:midas-contact@umich.edu).

For more, see [midas.umich.edu](http://midas.umich.edu)

## We gratefully thank our symposium sponsors:

Affiliate Program Sponsor:



Platinum Sponsors:




Gold Sponsor:



 ADVANCED RESEARCH COMPUTING			
COMPUTATIONAL SCIENCE	DATA SCIENCE	TECHNOLOGY SERVICES	CONSULTING SERVICES
MICDE	MIDAS	ARC-TS	CSCAR

MIDAS is a unit of Advanced Research Computing (ARC). ARC is part of the U-M Office of Research, and consists of MIDAS; the Michigan Institute for Computational Discovery and Engineering (MICDE); Advanced Research Computing - Technology Services (ARC-TS); and the Center for Statistical Consulting and Research (CSCAR).

Sign up for the ARC newsletter at: [arc.umich.edu/contact](http://arc.umich.edu/contact)