DATA ON A MISSION

INNOVATION AND APPLICATION @ COLUMBIA

Data Science Researchers
Demo and Present
New Work

March 31, 2015

LOW MEMORIAL LIBRARY, ROTUNDA
COLUMBIA UNIVERSITY

COLUMBIA UNIVERSITY
Data Science Institute

SUMMIT AGENDA

#DataOnAMission
@DSI_Columbia

COLUMBIA UNIVERSITY
Data Science Institute

On behalf of the Data Science Institute, welcome to our first ever Innovation and Application event. Today we have the opportunity to showcase data science research from across the university that is helping to transform society. Researchers will provide a hands-on look at the research and technologies they have developed, much of it with practical benefits for individuals, companies and cities, with relevance to the fields of finance, medicine and even sports.

You will have the opportunity to see Columbia's latest work in robotics and the Internet of things; Software that can quickly size up a product's carbon footprint and other environmental impacts; brain imaging techniques and computer models that are helping major league baseball teams identify the best batters and reassess their training methods.

You will also see smart building technology that is helping some of NYC's biggest landlords save millions in heating and cooling costs.

The Data Science Institute's Affiliate Program has been built upon three key tenets: our reputation, education and research. It's creation enables us to foster a partnerships between academia and industry. We want to establish strong connections so that our graduates are well prepared for the quickly changing tech workplace and so that we can also take good ideas developed at Columbia, and beyond academia, and bring them to market. To learn more about how your company can get involved, please see our website: https://industry.datascience.columbia.edu/industry-affiliates-program.

We hope you enjoy the experience and continue the journey with us to unlock the power of Data Science.

Sharon K. Sputz
Director of
Strategic Programs

WELCOME MESSAGE

TABLE OF CONTENTS

Registration and Breakfast 8:30AM – 9:25AM

Welcome and Introduction of Opening Keynote 9:25AM – 9:30AM

Kathleen R. McKeown, Director of the Data Science Institute and Henry and Gertrude Rothschild Professor of Computer Science

Opening Keynote Address 9:30AM – 10:05AM

TO BE ANNOUNCED

Lightning Talks: Session I 10:05AM – 10:45AM

SESSION CHAIR: *Tony Jebara, Associate Professor of Computer Science*

David Blei, Professor of Statistics and Computer Science
Yaniv Erlich, Assistant Professor of Computer Science
Andrew Gelman, Professor of Statistics and Political Science

Break 10:45AM – 11:00AM

Panel Session: Data from Sensing 11:00AM – 12:00PM

MODERATOR: *Patricia J. Culligan, Associate Director of the Data Science Institute and Professor of Civil Engineering and Engineering Mechanics*

PANELISTS:

Maria Feng, Rewick Professor of Civil Engineering
Henning G. Schulzrinne, Julian Clarence Levi Professor of Mathematical Methods and Computer Science and Professor of Electrical Engineering
Ioannis (John) Kymissis, Associate Professor of Electrical Engineering

Lightning Talks: Session II 12:00PM – 12:40PM

SESSION CHAIR: *Martin B. Haugh, Associate Professor of Professional Practice in the Department of Industrial Engineering and Operations Research*

Paul Glasserman, Jack R. Anderson Professor of Business
Garud Iyengar, Professor and Department Chair of Industrial Engineering and Operations Research
Sanat Kumar, Professor and Department Chair of Chemical Engineering

Pre-Lunch Introduction to Live Demos 12:40PM – 12:45PM

Networking Lunch 12:45PM – 2:00PM

Live Demos 2:00PM – 4:30PM

Lightning Talks: Session III 4:30PM – 5:10PM

SESSION CHAIR: *Steven M. Bellovin, Professor of Computer Science*

Augustin Chaintreau, Assistant Professor of Computer Science
Roxana Geambasu, Assistant Professor of Computer Science
Lance Weiler, Director of Experiential Learning and Applied Creativity
Susan McGregor, Assistant Professor of Journalism

Introduction of Closing Keynote 5:10PM – 5:15PM

Matthew C. Waxman, Liviu Librescu Professor of Law

Closing Keynote Address 5:15PM – 5:50PM

Andrew McLaughlin | Partner, Betaworks; CEO, Digg; CEO, Instapaper; Senior Fellow, School of International and Public Affairs, Columbia University

Closing Remarks 5:50PM – 5:55PM

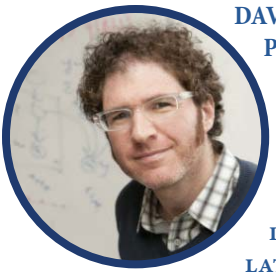
Sharon K. Sputz, Director of Strategic Programs for the Data Science Institute

Reception 5:55PM – 7:00PM

SESSION CHAIR: **TONY JEBARA**
ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

Tony Jebara is Associate Professor of Computer Science at Columbia University. He chairs the Center on Foundations of Data Science as well as directs the Columbia Machine Learning Laboratory. His research intersects computer science and statistics to develop new frameworks for learning from data with applications in social networks, spatio-temporal data, vision and text. Jebara has founded and advised several startups including Sense Networks (acquired by yp.com), AchieveMint, Agolo, Ufora, and Bookt (acquired by RealPage NASDAQ:RP). He has published over 100 peer-reviewed papers in conferences, workshops and journals including NIPS, ICML, UAI, COLT, JMLR, CVPR, ICCV, and AISTAT. He is the author of the book Machine Learning: Discriminative and Generative and co-inventor on multiple patents in vision, learning and spatio-temporal modeling. In 2004, Jebara was the recipient of the Career award from the National Science Foundation. His work was recognized with a best paper award at the 26th International Conference on Machine Learning, a best student paper award at the 20th International Conference on Machine Learning as well as an outstanding contribution award from the Pattern Recognition Society in 2001. Jebara's research has been featured on television (ABC, BBC, New York One, TechTV, etc.) as well as in the popular press (New York Times, Slash Dot, Wired, Businessweek, IEEE Spectrum, etc.). He obtained his PhD in 2002 from MIT. Esquire magazine named him one of their Best and Brightest of 2008. Jebara has taught machine learning to well over 1000 students (through real physical classes).

Jebara was a Program Chair for the 31st International Conference on Machine Learning (ICML) in 2014. Jebara was Action Editor for the Journal of Machine Learning Research from 2009 to 2013, Associate Editor of Machine Learning from 2007 to 2011 and Associate Editor of IEEE Transactions on Pattern Analysis and Machine Intelligence from 2010 to 2012. In 2006, he co-founded the NYAS Machine Learning Symposium and has served on its steering committee since then.



DAVID BLEI
PROFESSOR OF STATISTICS AND COMPUTER SCIENCE

"DATA SCIENCES FOR DOCUMENTS AND CLICKS"

Topic modeling algorithms analyze a document collection to estimate its latent thematic structure. However, many

COLLECTIONS CONTAIN AN ADDITIONAL TYPE OF DATA: HOW PEOPLE USE THE DOCUMENTS. FOR EXAMPLE, READERS CLICK ON ARTICLES IN A NEWSPAPER WEBSITE, SCIENTISTS PLACE ARTICLES IN THEIR PERSONAL LIBRARIES, AND LAWMAKERS VOTE ON A COLLECTION OF BILLS. BEHAVIOR DATA IS ESSENTIAL BOTH FOR MAKING PREDICTIONS ABOUT USERS (SUCH AS FOR A RECOMMENDATION SYSTEM) AND FOR UNDERSTANDING HOW A COLLECTION AND ITS USERS ARE ORGANIZED.

I WILL REVIEW THE BASICS OF TOPIC MODELING AND DESCRIBE OUR RECENT RESULTS ON COLLABORATIVE TOPIC MODELS, MODELS THAT SIMULTANEOUSLY ANALYZE A COLLECTION OF TEXTS AND ITS CORRESPONDING USER BEHAVIOR. WE STUDIED COLLABORATIVE TOPIC MODELS ON 80,000 SCIENTISTS' LIBRARIES FROM MENDELEY AND 100,000 USERS' CLICK DATA FROM THE ARXIV. COLLABORATIVE TOPIC MODELS ENABLE INTERPRETABLE RECOMMENDATION SYSTEMS, CAPTURING SCIENTISTS' PREFERENCES AND POINTING THEM TO ARTICLES OF INTEREST. FURTHER, THESE MODELS CAN ORGANIZE THE ARTICLES ACCORDING TO THE DISCOVERED PATTERNS OF READERSHIP. FOR EXAMPLE, WE CAN IDENTIFY ARTICLES THAT ARE IMPORTANT WITHIN A FIELD AND ARTICLES THAT TRANSCEND DISCIPLINARY BOUNDARIES.

David Blei joined Columbia in Fall 2014 as a Professor of Computer Science and Statistics. His research involves probabilistic topic models, Bayesian nonparametric methods, and approximate posterior inference. He works on a variety of applications, including text, images, music, social networks, user behavior, and scientific data.

Professor Blei earned his Bachelor's degree in Computer Science and Mathematics from Brown University (1997) and his PhD in Computer Science from the University of California, Berkeley (2004). Before arriving to Columbia, he was an Associate Professor of Computer Science at Princeton University. He has received several awards for his research, including a Sloan Fellowship (2010), Office of Naval Research Young Investigator Award (2011), Presidential Early Career Award for Scientists and Engineers (2011), and Blavatnik Faculty Award (2013).

YANIV ERLICH
ASSISTANT PROFESSOR OF COMPUTER SCIENCE

"GENETIC MEDIA"

NOTHING IN GENETIC MEDICINE MAKES SENSE EXCEPT IN THE LIGHT OF FAMILIAL RELATIONSHIPS. HOWEVER, THE RECRUITMENT OF LARGE COHORTS OF EXTENDED KINSHIPS IS BOTH LOGISTICALLY CHALLENGING AND COST PROHIBITIVE. HERE, WE PRESENT A NOVEL STRATEGY TO ADDRESS

THIS CHALLENGE: HARNESSING EXISTING, FREE, AND MASSIVE WEB 2.0 SOCIAL NETWORK RESOURCES TO TRACE TRAITS IN EXTREMELY LARGE FAMILIES. WE COLLECTED 43 MILLION PUBLIC PROFILES FROM GENI.COM, THE WORLD'S LARGEST GENEALOGY-DRIVEN SOCIAL NETWORK. USING THIS INFORMATION, WE CONSTRUCTED A SINGLE PEDIGREE OF OVER TEN MILLION INDIVIDUALS SPANNING MANY GENERATIONS UP TO THE 15TH CENTURY. WE ALSO EVALUATED THE QUALITY OF THE PEDIGREE CONSTRUCTION BY STANDARD GENETIC TOOLS. AS A PROOF OF CONCEPT, WE USED THIS MASSIVE DATASET TO DISSECT THE GENETIC ARCHITECTURE OF LONGEVITY. WE WILL ALSO DISCUSS HOW OTHER TRAITS CAN BE STUDIED USING THIS APPROACH.

Dr. Yaniv Erlich is Assistant Professor of Computer Science at Columbia University and Core Member at the New York Genome Center. Prior to this position, he was a Principal Investigator at the Whitehead Institute, MIT as the Andria and Paul Heafy Family Fellow. He received a bachelor's degree from Tel-Aviv University, Israel (2006) and a PhD from the Watson School of Biological Sciences at Cold Spring Harbor Laboratory (2010). Dr. Erlich's research interests are computational human genetics. Dr. Erlich is the recipient of the Burroughs Wellcome Career Award (2013), Harold M. Weintraub award (2010), the IEEE/ACM-CS HPC award (2008), and he was selected as one of 2010 Tomorrow's PIs team of Genome Technology.

ANDREW GELMAN
PROFESSOR OF STATISTICS AND POLITICAL SCIENCE

"STAN: A FLEXIBLE, OPEN-SOURCE PROGRAM FOR BAYESIAN DATA ANALYSIS

WE ARE DEVELOPING STAN, A C++ ENVIRONMENT FOR BAYESIAN INFERENCE THAT USES ADVANCED STATISTICAL AND NUMERICAL ALGORITHMS TO ALLOW USERS TO SPECIFY CUSTOM MODELS AND USE THEM FOR APPLIED RESEARCH AND DEVELOPMENT. WE ARE WORKING WITH COMPANIES INCLUDING NOVARTIS AND YOUGOV TO DEVELOP HIERARCHICAL MODELS FOR PHARMACOLOGY, POLITICAL POLLING, AND MANY OTHER APPLICATIONS. STAN IS FULLY FREE AND OPEN-SOURCE AND OUR WORK IS IN COLLABORATION WITH THE OPEN-SOURCE COMMUNITY.

Dr. Yaniv Erlich is Assistant Professor of Computer Science at Columbia University and Core Member at the New York Genome Center. Prior to this position, he was a Principal Investigator at the Whitehead Institute, MIT as the Andria and Paul Heafy Family Fellow. He received a bachelor's degree from Tel-Aviv University, Israel (2006) and a PhD from the Watson School of Biological Sciences at Cold Spring Harbor Laboratory (2010). Dr. Erlich's

research interests are computational human genetics. Dr. Erlich is the recipient of the Burroughs Wellcome Career Award (2013), Harold M. Weintraub award (2010), the IEEE/ACM-CS HPC award (2008), and he was selected as one of 2010 Tomorrow's PIs team of Genome Technology.

Andrew Gelman is a professor of statistics and political science and director of the Applied Statistics Center at Columbia University. He has received the Outstanding Statistical Application award from the American Statistical Association, the award for best article published in the American Political Science Review, and the Council of Presidents of Statistical Societies award for outstanding contributions by a person under the age of 40. His books include Bayesian Data Analysis (with John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Don Rubin), Teaching Statistics: A Bag of Tricks (with Deb Nolan), Data Analysis Using Regression and Multilevel/Hierarchical Models (with Jennifer Hill), Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do (with David Park, Boris Shor, and Jeronimo Cortina), and A Quantitative Tour of the Social Sciences (co-edited with Jeronimo Cortina).

Andrew has done research on a wide range of topics, including: why it is rational to vote; why campaign polls are so variable when elections are so predictable; why redistricting is good for democracy; reversals of death sentences; police stops in New York City, the statistical challenges of estimating small effects; the probability that your vote will be decisive; seats and votes in Congress; social network structure; arsenic in Bangladesh; radon in your basement; toxicology; medical imaging; and methods in surveys, experimental design, statistical inference, computation, and graphics.

MODERATOR: **PATRICIA J. CULLIGAN**
ASSOCIATE DIRECTOR, DATA SCIENCE INSTITUTE AND PROFESSOR OF CIVIL ENGINEERING AND ENGINEERING MECHANICS

A leader in the field of water resources and urban sustainability, Culligan has worked extensively with The Earth Institute's Urban Design Lab at Columbia University to explore novel, interdisciplinary solutions to the modern day challenges of urbanization, with a particular emphasis on the City of New York. Culligan is the director of a joint interdisciplinary Ph.D. program between Columbia Engineering and the Graduate School of Architecture Planning and Preservation that focuses on designs for future cities, including digital city scenarios. Her research group is active in investigating the opportunities for green infrastructure, social networks and advanced measurement and sensing technologies to improve urban water, energy, and environmental management.

Culligan received her M.Phil. and Ph.D. from the University of Cambridge, and was on the faculty at M.I.T before joining Columbia in 2003. She has received numerous awards for her contributions in engineering research and education, including the National Science Foundation's CAREER Award, the Egerton Career Development Chair, M.I.T's Arthur C. Smith Award for contributions to undergraduate life, Columbia Engineering School Alumni Association's Distinguished Faculty Award, and Columbia's Presidential Teaching Award.

Culligan serves on the National Academies Nuclear and Radiation Studies Board and the Board of Earth Sciences and Resources Committee on Geological and Geotechnical Engineering. In 2011, she was elected to the Board of Governors of the American Society of Civil Engineer's Geo-Institute. She is the author or co-author of six books, two book chapters, and over 70 referred scientific publications and 110 technical articles.

PANELIST: **MARIA FENG**
REWICK PROFESSOR OF CIVIL ENGINEERING

Professor Feng received her B.S. from Southeast University in China in 1982 and Ph.D. in mechanical engineering from the University of Tokyo in 1992. She started her academic career at Princeton University in 1990 as a Research Associate. She joined the faculty at the University of California, Irvine (UCI) in 1992 where he served at first as an Assistant Professor, then an Associate Professor,

a Professor, and most recently as a Henning Chancellor's Professor. She was the founding Director of the Center for Advanced Monitoring and Damage Inspection at UCI. Professor Feng joined the Columbia faculty in 2012 as Renwick Professor of Civil Engineering, an endowed professorship.

Professor Feng's research is on the forefront of multidisciplinary science and engineering in sensors, structural health monitoring, smart structure and system control for civil infrastructure and military applications, with an emphasis on structural safety and system resilience against natural and man-made hazards. She has made a number of original contributions to the state-of-the-art in both academic research and engineering practice through the development of novel fiber optic dynamic sensors, vision-based remote sensors, microwave imaging technology, vibration-based system identification algorithms for damage assessment, as well as the friction-controllable sliding isolation system and mega-sub structures for wind and seismic hazard mitigation.

Professor Feng's achievements have been recognized by her election as a Fellow of the American Society of Civil Engineers (ASCE) and numerous national and international awards. The honors include the CAREER Award by the National Science Foundation, the Collingwood Prize by ASCE, and the Alfred Noble Prize awarded jointly by the ASCE, the American Society of Mechanical Engineers, the Institute of Electrical and Electronics Engineers, the American Institute of Mining, Metallurgical, and Petroleum Engineers, and the Western Society of Engineers. Professor Feng also received the Water L. Huber Civil Engineering Research Prize from ASCE "for innovative, interdisciplinary and practical research on sensing, monitoring and controlling dynamic response of civil engineering systems subjected to earthquake and wind loads." In addition, she has received the Best Paper Award and the Best Presentation Award by the Japan Society of Instrument and Control Engineers, along with recognitions from other professional journals and conferences. Her work has been reported by national media, including a special feature "The Bridge Doctor" on NBC Nightly News with Brian Williams. She was named the Top Researcher on Wearable Sensors by MIT Technology Review.

PANELIST: HENNING G. SCHULZRINNE JULIAN CLARENCE LEVI PROFESSOR OF MATHEMATICAL METHODS AND COMPUTER SCIENCE AND PROFESSOR OF ELECTRICAL ENGINEERING

Professor Henning Schulzrinne, Levi Professor of Computer Science at Columbia University, received his Ph.D. from the University of Massachusetts in Amherst, Massachusetts. He was an MTS at AT&T Bell Laboratories and an associate department head at GMD-Fokus (Berlin), before joining the

Computer Science and Electrical Engineering departments at Columbia University. He served as chair of the Department of Computer Science from 2004 to 2009, as Engineering Fellow at the US Federal Communications Commission (FCC) in 2010 and 2011, and as Chief Technology Officer and Technical Advisor at the FCC from 2012.

Professor Schulzrinne has published more than 250 journal and conference papers, and more than 70 Internet RFCs. Protocols co-developed by him, such as RTP, RTSP and SIP, are now Internet standards, used by almost all Internet telephony and multimedia applications. His research interests include Internet multimedia systems, ubiquitous computing, and mobile systems.

He is a Fellow of the ACM and IEEE, has received the New York City Mayor's Award for Excellence in Science and Technology, the VON Pioneer Award, TCCC service award, IEEE Region 1 William Terry Award for Lifetime Distinguished Service to IEEE, the UMass Computer Science Outstanding Alumni recognition and is a member of the Internet Hall of Fame.

PANELIST: **IOANNIS (JOHN) KYMISSIS**
ASSOCIATE PROFESSOR OF ELECTRICAL ENGINEERING

Ioannis Kymissis joined the Electrical Engineering faculty in 2006. He teaches courses in solid state devices and display technology. He obtained his S.B., M.Eng. and Ph.D. degrees from MIT, and also participated in a cooperative program through which he completed his M.Eng. thesis at IBM Research. He also held a postdoctoral appointment at MIT and worked as a consulting engineer at QDVision before joining Columbia.

Professor Kymissis' research focuses on the fabrication, characterization, and applications of thin film electronics, with a particular focus in the applications of organic semiconductors and recrystallized silicon devices. In addition to his teaching and research work, he also serves as the Editor-in-Chief of the Journal of the Society for Information Display.

SESSION CHAIR: **MARTIN B. HAUGH**
ASSOCIATE PROFESSOR OF PROFESSIONAL PRACTICE IN THE DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH

Martin Haugh originally joined the Industrial Engineering and Operations Research Department in January 2002 after completing his PhD in Operations Research at MIT. He was a faculty member in the IEOR department until June 2005 and during this time his teaching and research focused on financial engineering. Between 2005 and 2009 he worked as a quant in the hedge fund industry in New York and London. He then returned to academia and the IEOR department in July 2009.

His current research interests include financial engineering, risk management, machine learning and Markov decision processes.

PAUL GLASSERMAN
JACK R. ANDERSON PROFESSOR OF BUSINESS

"FINANCIAL NETWORKS BASED ON PARTIAL DATA"

NETWORK MODELS HAVE BEEN PROPOSED AS TOOLS TO STUDY THE SPREAD OF SHOCKS THROUGH THE FINANCIAL SYSTEM AND THE CONSEQUENCES OF INTERCONNECTEDNESS. THE NODES IN THESE MODELS ARE FINANCIAL INSTITUTIONS AND THE LINKS BETWEEN THEM REFLECT PAYMENT OBLIGATIONS. IN PRACTICE, DETAILED INFORMATION ABOUT PAYMENT OBLIGATIONS IS UNAVAILABLE TO THE PUBLIC AND EVEN TO REGULATORS, SO APPLYING NETWORKS MODELS REQUIRES ADDRESSING THE PROBLEM OF PARTIAL DATA AVAILABILITY. APPROACHES TO THIS PROBLEM INCLUDE DEVELOPING BOUNDS BASED ON PARTIAL DATA, IMPUTING FIXED VALUES THROUGH OPTIMIZATION, AND SIMULATION CONSTRAINED BY PARTIAL DATA. REGULATORY DEVELOPMENTS HAVE EXPANDED THE SCOPE OF PARTIAL DATA AVAILABLE FOR USE WITH THESE METHODS.

Paul Glasserman is the Jack R. Anderson Professor of Business at Columbia Business School, where he serves as research director of the Program for Financial Studies. In 2011-2012, he was on leave from Columbia, working full-time at the Office of Financial Research in the U.S. Treasury Department, where he currently serves as a part-time consultant. His work with the OFR has included research on stress testing, financial networks, contingent capital, and counterparty risk. Paul's publications include the book Monte Carlo Methods in Financial Engineering (Springer, 2004), which received the 2006 Lanchester

Prize and the 2005 I-Sim Outstanding Publication Award. Glasserman is a past recipient of the National Young Investigator Award from the National Science Foundation (1994 - 99), IBM University Partnership Awards (1998 - 2001), the TIMS Outstanding Simulation Publication Award (1992), and the Erlang Prize in applied probability (1996). He was named Risk magazine's Quant of the Year in 2007. Paul served as senior vice dean of Columbia Business School in 2004-2008 and was interim director of its Sanford C. Bernstein Center for Leadership and Ethics in 2005-2007

GARUD IYENGAR
PROFESSOR AND DEPARTMENT CHAIR OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH

"EXAMPLES OF DATA-DRIVEN OPERATIONS: FROM ADVERTISING TO SOLAR MICRO-GRIDS"

OPERATIONS RESEARCH IS CONCERNED WITH CHOOSING DECISIONS THAT MINIMIZE AN APPROPRIATELY DEFINED METRIC OF OPERATING COSTS. THIS REQUIRES ESTIMATING THE DISTRIBUTIONS OVER FUTURE SCENARIOS. SO, A NATURAL RECIPE FOR DATA-DRIVEN OPERATIONS IS TO USE MACHINE LEARNING TO ESTIMATE THE SCENARIO DISTRIBUTION, AND USE THE ESTIMATED DISTRIBUTION COMPUTE THE OPTIMAL DECISION. HOWEVER, ONE CAN DO MUCH BETTER BY ESTIMATING THE DISTRIBUTION AND THE DECISION TOGETHER! WE WILL DISCUSS EXAMPLES OF THIS APPROACH IN SEVERAL DIFFERENT CONTEXTS.

Professor Garud Iyengar joined Columbia University's Industrial Engineering and Operations Research Department in 1998. Professor Iyengar teaches courses in simulation and optimization.

Professor Iyengar's research interests include convex optimization, robust optimization, mathematical and computational finance, machine learning and computational biology.

SANAT KUMAR
PROFESSOR AND DEPARTMENT CHAIR OF CHEMICAL ENGINEERING

"MATERIALS DISCOVERY"

AGOLO

Sage Wohns, Mohamed AlTantawy, Will Viet, Julian Norton, and Jason Choo

Agolo helps big data consumers and content providers by generating personalized summaries at machine scale. Everyday, our natural language processing engine produces over 63,000 summaries, generated from analyzing over 3.9M tweets and 190k links a day. Agolo creates summaries from unstructured data sources like the news, social media, and market data. Our natural language processing engine listens to real-time feeds in order to cluster related information. Agolo transforms these clusters of unstructured data into readable summaries. Our summarization engine is built to capture the natural progression of a story unfolding in real-time. We identify multiple voices for the same story expressed in different formats.

ANTIVIRUS IN SILICON

John Demme and Adrian Tang

Anti-virus (AV) software is fundamentally broken. AV systems today rely on correct functioning of not only the AV software but also the underlying OS and VMM. Thus proper functioning of software AV requires millions of lines of complex code – which houses thousands of bugs – to work correctly. Needless to say, and as evidenced in numerous software AV attacks, effective software AV systems have been difficult to build. At the same time, malware incidents are increasing and there is strong demand for good anti-virus solutions; the software anti-virus market is estimated at close to 8B dollars annually.

In this work we present a new class of robust AV systems called Silicon anti-virus systems. Unlike software AV systems, these systems are lean and mostly implemented in hardware to avoid reliance on complex software, but, like software AV systems, are updatable in the field when new malware is encountered. We describe the first generation of silicon AV that uses simple machine learning techniques with existing performance counter infrastructure. These systems form a very effective first-line, energy- efficient defense against malware.

AUGMENTED REALITY FOR TASK ASSISTANCE

Steve Feiner, Barbara Tversky, Carmine Elvezio, Ohan Oda, and Mengu Sukan

We will demonstrate our research on augmented reality user interfaces that support collaboration between two users at physically distant sites. We are developing and evaluating computer-aided 3D referencing techniques that allow a remote instructor to provide instructions to a local trainee, to assist the trainee in performing tasks at their site. These tasks involve positioning and orienting physical objects relative to each other. The instructor and trainee each wear a stereoscopic, head-tracked, see-through, head-worn display that creates an augmented reality view in which 3D virtual graphics are overlaid interactively on the user's view of the real world. The instructor can switch between observing the trainee's site through live stereo video captured by cameras on the trainee's

head-worn display, or through 3D models that dynamically track the physical objects manipulated by the trainee. Attendees will be able to try performing tasks as either the instructor or trainee.

We will also demonstrate our research on ParaFrustum, a way to represent the set of strategic viewing positions and orientations from which a task can be accomplished, so that the user can be efficiently guided to assume a viewing position and orientation from this set, Attendees will experience some of the augmented reality visualizations that we have developed to communicate this information to a user wearing a stereoscopic, head-tracked, see-through, head-worn display.

CITIBIKE REBALANCING STUDY

Juan Francisco Saldarriaga

As has been recently documented by the press, one of the major challenges that Citi Bike is facing is the rebalancing of their stations. As origins and destinations of Citi Bike trips are not necessarily symmetrical during the day, Citi Bike has been forced to constantly move bikes around the city, taking them from full stations and delivering them to empty ones. This problem is both financially expensive and frustrating for Citi Bike users: many people complain about either not finding bikes at their stations of origin or not finding empty spots when they arrive at their final destinations. To study this problem we have created a series of visualizations which should serve as a starting point for further analysis.

COCLEAR: MARRYING DATA SCIENCE AND ENVIRONMENTAL ENGINEERING TO LEAP FROM SUSTAINABLE PRODUCTS TO COMPETITIVE ENTERPRISES

Christoph Meinrenken, Daniel Chen, Sally Paridis and Erika Whillas

Product life cycle assessment (LCA) offers tremendous opportunities to improve consumer products' competitiveness as well as reduce their environmental impact. However, carrying out LCA at the scale of large companies has been impossible, especially since large portions of a product's life cycle costs and environmental impacts occur outside the manufacturing company's immediate control – namely up and downstream in the typically global product supply chain. CoClear, an NY startup of Columbia alums, uses patented patented data mining and machine learning algorithm to dramatically increase the speed of LCA and pairs this with state of the art interactive visualizations and mapping tools.

DEMONSTRATION OF COMPRESSED SAMPLING FOR RAPID RF SPECTRUM SCANNING

Peter Kinget, John Wright, Tanbir Haque, Rabia Tugce Yazicigil and Jeffrey Yuan

Wireless systems have become an essential part of every sector of the global economy. In addition to the existing commercial systems including mobile cellular,

WiFi communications and global positioning system, emerging systems like video over wireless, Internet of Things and machine-to-machine communications are expected to increase the mobile wireless data traffic by several orders of magnitude over the coming decades while the natural resources like energy and radio spectrum remain scarce. If the consumer demand for instantaneous, over the air access to large volume of content continues to grow at its current rate in the US, a 500MHz to 1GHz spectrum deficit is expected in the near to long term.

Cognitive radio is a paradigm proposed to overcome the existing challenge of underutilized spectrum. Future cognitive radio systems employing multi-tiered, shared spectrum access (MTSSA) are expected to deliver superior spectrum efficiency over current scheduled access systems. Lower tiered 'smart' devices opportunistically use the underutilized spectrum and will need spectrum sensing for incumbent detection and interferer avoidance. Integrated interference detectors need to be fast, wideband and energy efficient while only requiring moderate sensitivity. It is envisioned that the future receivers will use interference detectors to quickly detect the presence of a few interferers in a wideband spectrum with fine resolution during the designated slot boundaries and will be reconfigured on a frame or a slot basis. Compressed sensing as a blind sub-Nyquist sampling method has the potential to deliver this energy efficient and rapid interferer detection over a wideband spectrum. Compressed sensing allows users to gather data and information about the spectrum by sampling at a rate defined by information content rather than instantaneous bandwidth of the spectrum.

We demonstrate a rapid interferer detector exploiting compressed sampling (CS) with a quadrature analog-to-information converter (QAIC)*. Our architecture is two orders of magnitude more energy efficient compared to traditional spectrum scanners and one order of magnitude more energy efficient compared to existing CS spectrum sensors. To scan a wideband 1GHz frequency span with a 20MHz resolution bandwidth, it offers 50x faster scan time compared to traditional sweeping spectrum scanners and 6.3x compressed aggregate sampling rate compared to traditional concurrent Nyquist rate spectrum scanners.

*R.T. Yazicigil, T. Haque, M.R. Whalen, J. Yuan, J. Wright, and P.R. Kinget, "A 2.7-3.7GHz Rapid Interferer Detector Exploiting Compressed Sampling with a Quadrature Analog-to-Information Converter," IEEE International Solid-State Circuits Conference (ISSCC), Feb. 2015.

DI-BOSS™

Roger Anderson, Francesco Calabro, Albert Boulanger, Ashish Gagneja, Leon Wu,Hooshmand Shokri Razaghi, and Promiti Dutta

Di-BOSS™ is the world's first Digital Building Operating System for buildings. All sub-systems are integrated into a Systems Integration Facility that archives past data and uses machine learning algorithms to predict future forecasts via the Total Property Optimizer built and maintained by Columbia University's Center

for Computational Learning Systems.

ENABLING MM-WAVE SAME-CHANNEL FULL-DUPLEX WIRELESS COMMUNICATION

Harish Krishnaswamy and Tolga Dinc

Same-channel full-duplex where the transmitter and the receiver operate at the same time and at the same frequency is an exciting emergent technology. It can double the spectral efficiency and data throughput. Recently, we have developed the world's first fully-integrated transceiver (45nm SOI CMOS 60 GHz direct-conversion) for same-channel full-duplex wireless communication. By using our transceiver, we will demonstrate a mm-wave same-channel full-duplex wireless link.

ENERGY HARVESTING ACTIVE NETWORKED TAGS (ENHANTS)

Peter Kinget, Ioannis Kymissis, Dan Rubenstein, Gil Zussman and Robert Margolies

In this project we are developing Energy-Harvesting Active Networked Tags (EnHANTs). EnHANTs are small, flexible, and energetically self-reliant devices that can be attached to objects that are traditionally not networked (e.g., books, furniture, walls, doors, toys, keys, clothing, and produce), thereby providing the infrastructure for various novel tracking applications. Examples of these applications include locating misplaced items, continuous monitoring of objects (items in a store, boxes in transit), and determining locations of disaster survivors.

Recent advances in ultra-low-power wireless communications, ultra-wideband (UWB) circuit design, and organic electronic harvesting techniques will enable the realization of EnHANTs in the near future. In order for EnHANTs to rely on harvested energy, they have to spend significantly less energy than Bluetooth, Zigbee, and IEEE 802.15.4a devices. Moreover, the harvesting components and the ultra-low-power physical layer have special characteristics whose implications on the higher layers have yet to be studied (e.g., when using ultra-low-power circuits, the energy required to receive a bit is significantly higher than the energy required to transmit a bit).

We have been developing the EnHANTs prototypes and testbed over the last 4 years. The prototypes communicate using in-house developed ultra-wide-band impulse radio (UWB-IR) transceivers. They create a multihop network and adapt the topology and communication patterns to the light energy harvested by the solar cells.

In this demo, we present our EnHANTs prototypes and testbed. In our interactive demo we show several EnHANT prototypes communicating with each other. The prototypes determine how much power their solar cells are generating, and how much energy their EHM battery contains. Based on that, the prototypes jointly determine their communication parameters (data rates, sleep/wake cycles,

network topology).

Using a graphical monitoring system, demo participants can quickly observe the changes in the communication parameters. Demo participants can observe the network adapt to varying light energy conditions using a user-operated module of the light system. Overall, using our small scale testbed of four EnHANTs prototypes, we show novel functionalities across the layers of the protocol stack, from UWB-IR communications to energy harvesting-adaptive networking protocols.

GEWORKBENCH: A PLATFORM FOR INTEGRATED GENOMICS

Aris Floratos and Kenneth Smith

geWorkbench (genomics Workbench) is an open source Java desktop application that provides access to an integrated suite of tools for the analysis and visualization of data from a wide range of genomics domains including gene expression, sequence, protein structure and systems biology (http://www.geworkbench.org). More than 70 distinct plug-in modules are currently available implementing both classical analyses (several variants of clustering, classification, Gene Ontology, homology detection, etc.) as well as state-of-the-art algorithms for the reverse-engineering of regulatory networks and for protein structure prediction. geWorkbench provides seamless access to many public genomic databases and annotation sources and to sophisticated remote computational infrastructure, including a 6000 CPU high-performance computing cluster at the Department of Systems Biology. Through geWorkbench researchers can gain access to unique biological Big Data collections assembled by investigators at Columbia University, including a growing repository of computationally and experimentally derived gene interaction networks for a large number of cellular contexts (the Cellular Network Knowledge Base, CNKB). Attention has been paid to the development of visually rich graphical user interfaces designed to facilitate the interactive exploration of multidimensional genomic data and analysis results.

geWorkbench-web is a web-enabled version of geWorkbench, developed to eliminate the burden of local software installation and to provide instant, browser-based access to analysis and visualization components. Currently in beta release (http://geworkbench.c2b2.columbia.edu/geworkbench/), geWorkbench-web already incorporates many of the most powerful features of the desktop version including tools for the reverse-engineering of cellular regulatory network, differential gene expression analysis, ANOVA, hierarchical clustering, and queries against the CNKB.

Both varieties of geWorkbench have been developed under the auspices of MAGNet, the National Center for the Multi-scale Analysis of Genomic and Cellular Networks, a large collaborative project involving faculty from academic Departments inside and outside Columbia University (http://magnet.c2b2.columbia.edu). MAGNet is one of eight National Centers for Biomedical Computing program (http://www.ncbcs.org/), a network of Centers tasked with

building the computational infrastructure for biomedical computing in the nation. geWorkbench was developed to integrate software developed through MAGNet, along with other tools and data sources, and make it available to biomedical researchers. geWorkbench has more than 2000 users.

GSPIN: A GMAIL PLUGIN FOR DEEP ANALYTICS OF EMAIL INTERACTIONS

Owen Rambow, Vinodkumar Prabhakaran and Michael Saltzman

The gSPIN system is a secure Gmail plugin that brings to its user the power of deep analytics of his/her own email interactions. The technology behind this tool is built over the past four years as part of the "Social Power in INteractions (SPIN)" project at the Center for Computational Learning Systems at Columbia University. The SPIN project uses advanced natural language processing and machine learning techniques to perform deep analytics on social interactions to understand the dialog patterns in them and reveal the underlying social power relations that exist between the interactants.

HUMAN-IN-THE-LOOP-GRASP PLANNING FOR ASSISTIVE MANIPULATION USING A NOVEL SEMG INTERFACE

Peter Allen, Yinxiao Li, and Jon Weisz

In this work, we present a prototype assistive robotic manipulation system that is accessible to individuals with severe motor impairments. By pairing state of the art real time grasp analysis with an augmented reality interface that is controlled through a small set of inputs, we can allow a severely impaired individual to grasp objects in multi-object, cluttered scenes. The interface that we present in this work uses a single electrode recording site to acquire muscle signals from behind the ear of the user. These muscles make an interesting target for assistive interfaces for a number of reasons. Control over these muscles is preserved even in patients with severe spinal injuries because the nerves that innervate them do not enter the spinal cord. These muscles are not directly involved in most other activities, and by carefully selecting the recording site, we can isolate the interaction with our interface from interfering with normal activities such as eating

INPATIENT MYNYP – ENGAGING HOSPITAL PATIENTS IN THEIR CARE

Suzanne Bakken, Steve Feiner, George Hripcsak, Susan Restaino, David Vawdrey, Ruth Masterson Creber, Beatriz Ryan, Jungmi Han and Jennifer Prey

Involving patients in their healthcare is a growing trend. Technology can provide consumers with direct access to their clinical health information. Such information is seldom available to patients in the hospital setting. A new project, Inpatient myNYP, is a personalized, dynamic patient portal that was created

by the Department of Biomedical Informatics in partnership with NewYork-Presbyterian Hospital (NYP). It provides patients with immediate access to their hospital records, which are directly sourced from NYP's electronic health record (EHR). Patients use Inpatient myNYP to view the members of their care team, track current and historical vital signs and hospital medications, review their documented hospital medications and allergies, document pain scores, record personal notes, and send questions and comments to their care team members. These questions and comments are made directly available to clinicians in the EHR.

MOBILE DIABETES DETECTIVE (MODD): HEALTH INFORMATION TECHNOLOGY FOR FACILITATING PROBLEM SOLVING IN DIABETES SELF-MANAGEMENT

George Hripcsak, Rita Kukafka, Arlene Smaldone, Lena Mamykina and Elizabeth Heitkemper

Well-developed problem solving is essential to successful diabetes management, and leads to improved self-management behaviors and glycemic control. Yet few informatics interventions specifically focus on helping individuals with diabetes engage in problem-solving and develop problem-solving skills. Mobile Diabetes Detective (MoDD) is a novel electronic tool that helps individuals with diabetes monitor their blood glucose levels, identify times of day when these levels are systematically higher or lower than the recommended ranges, identify daily behavior that may contribute to these problematic patterns, and choose alternative healthy behaviors to improve glycemic control.

NEURAL CORRELATES OF RAPID DECISION MAKING: A REAL-TIME SYSTEM FOR ASSESSING AND TRAINING BASEBALL HITTERS

Paul Sajda and Jordan Muraskin

In baseball, a hitter has a fraction of a second to decide whether the pitch will be a ball or a strike and to decide whether to swing at the pitch. One element of the rapid decision making process is determining what type of pitch is thrown, e.g., a fastball, curveball, or a slider, because the type of pitch constrains the potential trajectories of the ball. We have developed a real-time on-line demonstration system for tracking neural correlates of pitch recognition; the goal being to using these neural correlates to assess batter performance. The demo uses a 64-channel EEG amplifier and headset, a stimulus computer (to show simulated baseball pitches), an analysis computer, and one pre-trained user. After setting up the EEG cap on the user, the real-time neuro-tracking paradigm will begin. The user plays a simulated baseball game in which he/she has to decide whether to swing (press a button on the keyboard) or not swing (lack of keyboard button response) at the incoming pitch given a pre-pitch cue. While he/she is making this rapid decision in less than 600ms, we process the EEG data and find the spatial location of the pitch when he/she decided to swing or not swing. We then display this location

back to the user. At the end of the session, we are able to compile a complete baseball specific neuro-profile. This technology is being commercialized by the start-up company deCervo LLC (decervo.com) that is a spin-out from the LIINC lab.

NEWS ROVER

Shih-Fu Chang, Joseph G. Ellis, Brendan Jou, Hongzhi Li, Daniel Morozoff-Abezgauz

Columbia NewsRover system demonstrates novel multimodal analysis technologies for extracting important information from video, speech, and text in real time over 100 broadcast TV channels. It links such information over time and diverse sources in order to generate intuitive summaries addressing the key questions of Who, What, Where, When, and Why for real world events. It accomplishes the objectives by applying advanced multimedia and machine learning technologies developed in house at the Digital Video and Multimedia Lab. NewsRover capabilities provide the opportunities for developing next-generation information systems such as Personalized TV News and Augmented Information Workbench for consumers and analysts.

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS (OHDSI)

Adler Perotte and Joseph Romano

OHDSI is an open, multi-stakeholder, interdisciplinary collaborative whose goal is to create and apply open-source large-scale data analytic solutions that help extract the value of observational health data to improve human health and well-being. Functionally, OHDSI aims to release a number of publicly available, open-source data analytics and visualization tools, ranging in scope from patient-level prediction to population-level estimation. By computing statistics locally on health databases and only making the summary statistics publicly accessible, OHDSI software and resultant data summaries can remain open without compromising the confidentiality of individual patient health records. ACHILLES – the first of these tools to be released – is available as a demo on the OHDSI website. ACHILLES reports cumulative statistics on large clinical databases and presents them as visualizations arranged in a dashboard-style layout. As of August 2014, the network of OHDSI collaborators consisted of 70+ academic, industrial, and regulatory researchers, a number which continues to grow. OHDSI's future vision includes coverage of 1,000,000,000+ patient records for the generation of high-resolution worldwide health statistics.

PLACEMETER: A COLLABORATIVE MODEL FOR MODELING URBAN BEHAVIOR

David Fine and Jason Novack

Enabling agile urbanism with prolific, long-term data generated by simple sensors

and cloud processing.

PREJU – THE PREDICTION JUSTIFIER

Kathleen McKeown and Or Biran

PreJu is a software library that produces textual justifications of predictions made by third-party Machine Learning models. We will demonstrate how it works with a stock price prediction model.

RECOGNITION, REGRASPING, AND UNFOLDING OF DEFORMABLE OBJECTS USING PREDICTIVE THIN SHELL MODELING

Peter Allen, Yinxiao Li, and Jon Weisz

Deformable objects such as garments are highly unstructured, making them difficult to recognize and manipulate. In this paper, we propose a novel method to teach a two-arm robot to efficiently track the states of a garment from an unknown state to a known state by iterative regrasping. The problem is formulated as a constrained weighted evaluation metric for evaluating the two desired grasping points during regrasping, which can also be used for a convergence criterion. The result is then adopted as an estimation to initialize a regrasping, which is then considered as a new state for evaluation. The process stops when the predicted thin shell conclusively agrees with reconstruction. We show experimental results for regrasping a number of different garments including sweater, knitwear, pants, and leggings, etc.

REDOWL ANALYTICS – INSIDER THREAT ANALYTICS

David Pogemiller and Mauricio Renzi

RedOwl's Reveal software was built to address three critical challenges for security and compliance teams as they increasingly focus their areas on detecting and addressing human risk inside the institution:

– Create holistic visibility on internal employee activity, behaviors, and relationships across all forms of critical data (in a rapidly evolving data environment);

– Move from reactive to proactive in detecting (and mitigating) high-risk individuals, relationships and events;

– Enhance investigative response to reporting and alerting through improved context, increased speed of "decision", reducing false positives, and exposing a higher-level of scenarios to monitor against.

SECE: SENSE EVERYTHING, CONTROL EVERYTHING

Henning G. Schulzrinne and Jan Janak

SECE is an architectural framework for programmable heterogeneous smart object networks. A programmable smart object network can be used as a building

block in Cyber-Physical Systems (CPS) or Internet-of-Things (IoT) applications. SECE provides a distributed end-to-end programming model. An application can be deployed close to the devices it interacts with, e.g., to an application server in a local network. Portions of the application can be deployed directly onto the end devices (smart objects) at run-time, using the device's network programming APIs. The framework is based on standard open protocols, e.g., IPv6, HTTP, and RTP. The distributed nature of SECE applications leads to resilient systems without a central bottleneck. The end-to-end programmability feature makes it possible to implement latency-sensitive closed loop scenarios, commonly found in CPS applications.

SOCIAL NETWORK EXTRACTION FROM TEXT

Owen Rambow, Apoorv Agarwal and Ari Fay

Language is the primary tool that people use for establishing, maintaining and expressing social relations. This makes language the real carrier of social networks. We demonstrate our system (SINNET) that automatically extracts a social network from raw texts such as nineteenth century British literature, film screenplays, and other types of text that may be entered by the user of our system.

SOLESOUND: A WEARABLE DEVICE FOR COST-EFFECTIVE, LONG-TERM, OUT-OF-THE LAB, ACCURATE GAIT ANALYSIS

Sunil Agrawal, Damiano Zanotto, Emily Boggs, Minh Nguyen and Stephan Stansfield

Quantitative gait analysis enables clinicians to diagnose a number of neuromuscular disorders and to evaluate patient mobility before and after a surgery. Because it is conducted with camera-based motion capture systems and instrumented walkway devices, patients who lack access to gait laboratories or suffer from episodic gait impairments that are difficult to elicit in the laboratory settings cannot benefit from quantitative gait analysis.

The use of instrumented footwear that unobtrusively performs long-term, out of the lab measurements on subject's walking pattern is a promising way to overcome these limitations. Besides serving as assessment tools, such devices can also act as retraining tools that help regulate a patient's gait by means of acoustic or vibrotactile stimuli.

At the Columbia University Robotics and Rehabilitation (ROAR) Lab, we have developed SoleSound, a fully portable instrumented footwear that can accurately measure a rich set of spatiotemporal gait parameters and deliver action-related audio-tactile feedback in response to measured parameters to help the user regulate her gait. Unlike similar devices, SoleSound does not require a host computer. Most importantly, it can measure clinically relevant gait parameters - including inter-limb variables such as the base of walking - and synthesize continuous audio-tactile feedback based on these. These unique features enable SoleSound to quantify the immediate and long-term effects of gait retraining with

audio-tactile feedback without the need for laboratory equipment.

SUMMARIZING TOPICS: FROM WORD LISTS TO DESCRIPTIVE PHRASES

Lauren A. Hannah and Hanna Wallach

We introduce a statistically principled, computationally efficient, two-stage method for generating phrase-based topic summaries from the inferred parameters of any statistical topic model based on latent Dirichlet allocation. This method involves 1) identifying n-gram phrases and 2) selecting descriptive words and phrases for each topic using a novel metric, KALE, that balances distinctiveness and recognizability. We describe three different phrase-finding algorithms, including a new Bayesian algorithm, which does not rely on topic model parameters and therefore constitutes an general-purpose, stand-alone contribution to the phrase-finding literature. We provide a human-subjects evaluation of our two-stage topic summarization method, comparing summaries produced using each phrase-finding algorithm to summaries consisting of the most probable words for each topic. We also compare summaries produced by the best performing variant of our method to those produced by other commonly used topic summarization methods.

THE SYNAPSE

Laura Kurgan, Jochen Hartmann and Madeeha Merchant

The Synapse is an initiative for the communication of science at the Mortimer B. Zuckerman Mind Brain Behavior Institute in the Jerome L. Greene Science Center at Columbia University. The design for the space is being developed as a joint initiative between the Spatial Information Design Lab (Graduate School of Architecture, Planning and Preservation) and the Brown Institute for Media Innovation (Graduate School of Journalism).

Our design for the the space includes exhibits, visualizations and story-telling components, and draws upon many disciplines such as architecture, journalism and data science. The goal is to create a networked public space that will actively engage, and be genuinely accessible to, a heterogeneous population of visitors, including resident scientists and community members alike.

To develop aspects of the Synapse content, our team has began to actively pursue research into the 3d visualization and eventual printing of neurons, lead by Madeeha Merchant. So far we have acquired 30,000 different neurons from various sources and have examined 30 in close detail. We have worked with Columbia neuroscientists Randy Bruno, Raffa Yuste and Tom Jessell who have provided us with image stacks produced with a confocal microscope and we have then developed our own workflows to construct three-dimensional meshes based off this data.

WEARABLE DYNAMICALLY MODULATED SPINE BRACE

Sunil Agrawal, Joon Park and Paul Stegall

Adolescent idiopathic scoliosis affects 1-3% of the population in the United States. The current treatment methods require patients to wear a ridged plastic brace for 16-20 hours per day. The discomfort and social factors contribute to poor compliance with braces. The proposed robotic brace design utilizes two Stewart-Gough platforms in series, each platform controlled independently, either in position or force modes. The brace can control the motion or force in different regions of the spine to provide correction. The brace can monitor motion and force data in real-time to modify the course of treatment. It is hoped that the active robotic brace will lead to new treatment methods to improve brace outcomes. Human studies are currently being planned with this brace to characterize the stiffness of the human body and treatment of scoliosis patients. The brace has other human applications to provide rigidity to the human body during complex and heavy lifting tasks.

	TIMING
LOW LIBRARY	
• CitiBike Rebalancing Study	2:00PM
• CoClear	2:40PM
• Di-BOSS™	3:15PM
• geWorkbench	3:45PM
• News Rover	
• Placemeter	
• The Synapse	
NWC 10–I (NORTHWEST CORNER)	
• AGOLO	2:00PM
• gSPIN	2:40PM
• Inpatient myNYP	3:15PM
• Mobile Diabetes Detective (MoDD)	3:45PM
• Observational Health Data Sciences and Informatics (OHDSI)	
• PreJu	
• RedOwl Analytics	
• Social Network Extraction from Text	
• Summarizing Topics	
NWC 10–II (NORTHWEST CORNER)	
• Antivirus in Silicon	2:00PM
• Demonstration of Compressed Sampling for Rapid RF Spectrum Scanning	2:40PM
• Enabling mm-Wave Same-Channel Full-Duplex Wireless Communication	3:15PM
• Neural Correlates of Rapid Decision Making	3:45PM
• SoleSound	
• Wearable Dynamically Modulated Spine Brace	
CEPSR 6–8	
• Augmented Reality for Task Assistance *(6th fl)*	2:00PM
• Energy Harvesting Active Networked Tags (EnHANTs) *(8th fl)*	2:40PM
• Human-In-The-Loop-Grasp Planning for Assistive Manipulation Using a Novel sEMG Interface *(6th fl)*	3:15PM
• Recognition, Regrasping, and Unfolding of Deformable Objects Using Predictive Thin Shell Modeling *(6th fl)*	3:45PM
• SECE: Sense Everything, Control Everything *(7th fl)*	

SUGGESTED ATTENDEE ASSIGNED CLUSTERS

RED GROUP	TIME	SITE
APPROXIMATELY 30 MINUTES IN EACH LOCATION	2:00 PM	LOW
	2:40 PM	CEPSR 6–8
	3:15 PM	NWC 10–I
	3:45 PM	NWC 10–I

YELLOW GROUP	TIME	SITE
APPROXIMATELY 30 MINUTES IN EACH LOCATION	2:00 PM	NWC 10–I
	2:40 PM	NWC 10–II
	3:15 PM	CEPSR 6–8
	3:45 PM	LOW

BLUE GROUP	TIME	SITE
APPROXIMATELY 30 MINUTES IN EACH LOCATION	2:00 PM	NWC 10–II
	2:40 PM	NWC 10–I
	3:15 PM	LOW
	3:45 PM	CEPSR 6–8

GREEN GROUP	TIME	SITE
APPROXIMATELY 30 MINUTES IN EACH LOCATION	2:10 PM	CEPSR 6–8
	2:40 PM	LOW
	3:45 PM	NWC 10–II
	3:45PM	NWC 10–I

SESSION CHAIR: **STEVEN M. BELLOVIN**
PROFESSOR OF COMPUTER SCIENCE

Steven M. Bellovin is the Percy K. and Vidal L. W. Hudson Professor of Computer Science at Columbia University, where he does research on networks, security, and especially why the two don't get along, as well as related public policy issues. In his spare professional time, he does some work on the history of cryptography. He joined the faculty in 2005 after many years at Bell Labs and AT&T Labs Research, where he was an AT&T Fellow. He received a BA degree from Columbia University, and an MS and PhD in Computer Science from the University of North Carolina at Chapel Hill. While a graduate student, he helped create Netnews; for this, he and the other perpetrators were given the 1995 Usenix Lifetime Achievement Award (The Flame). Bellovin has served as Chief Technologist of the Federal Trade Commission. He is a member of the National Academy of Engineering and is serving on the Computer Science and Telecommunications Board of the National Academies, the Department of Homeland Security's Science and Technology Advisory Committee, and the Technical Guidelines Development Committee of the Election Assistance Commission; he has also received the 2007 NIST/NSA National Computer Systems Security Award and has been elected to the Cybersecurity Hall of Fame.

Bellovin is the co-author of Firewalls and Internet Security: Repelling the Wily Hacker, and holds a number of patents on cryptographic and network protocols. He has served on many National Research Council study committees, including those on information systems trustworthiness, the privacy implications of authentication technologies, and cybersecurity research needs; he was also a member of the information technology subcommittee of an NRC study group on science versus terrorism. He was a member of the Internet Architecture Board from 1996-2002; he was co-director of the Security Area of the IETF from 2002 through 2004.

AUGUSTIN CHAINTREAU
ASSISTANT PROFESSOR OF COMPUTER SCIENCE AND
ROXANA GEAMBASU
ASSISTANT PROFESSOR OF COMPUTER SCIENCE

"XRAY: NEW TOOL TO INCREASE TRANSPARENCY IN DATA DRIVEN SYSTEMS"

Today's Web services – including Google, Amazon, and Facebook – leverage user data for personalizing recommendations, targeting advertisements, and adjusting prices. Users currently have little insight, and at best coarse information, to monitor how and for which purposes their data are being used. What if we could tell

EXACTLY WHICH ITEM - WHETHER AN EMAIL YOU WROTE, A SEARCH YOU MADE, OR A WEBPAGE YOU VISIT - IS BEING USED TO DECIDE ON A TARGETED AD OR A RECOMMENDED PRODUCT FOR YOU? BUT HOW CAN WE TRACK DATA IN AN ENVIRONMENT WE DO NOT CONTROL?

IN THIS TALK, WE ARGUE THAT WITHOUT WEB TRANSPARENCY THE EXCITING WORLD OPEN WITH YOUR DATA THREATENS TO BECOME A BREEDING GROUND FOR DATA MISUSE, PRIVACY NEGLIGENCE, OR EVEN UNFAIR AND PREDATORY PRACTICES, DISCRIMINATING THE MOST VULNERABLE. AS A FIRST STEP TOWARD RESTORING WEB TRANSPARENCY, WE HAVE BUILT XRAY, THE FIRST FINE-GRAINED, ROBUST, AND SCALABLE TRACKING SYSTEM FOR PERSONAL DATA THE WEB. XRAY DIAGNOSES WHICH CLUE (I.E. EMAILS, VIEWED PRODUCTS) IS BEING USED AS TRIGGER TO WHICH OUTPUTS (I.E. TARGETED ADS, RECOMMENDED PRODUCTS, OR DIFFERENTIATED PRICES). XRAY IS SERVICE AGNOSTIC, EASY TO INSTANTIATE, AND LEVERAGE A NOVEL AND SIMPLE MECHANISM THAT, SURPRISINGLY AT FIRST, SHOWS THAT AS DATA IN OUR WEB PROFILE EXPANDS, THE AMOUNT OF RESOURCE REQUIRED FOR TRANSPARENCY GROWS ONLY LOGARITHMICALLY.

Augustin Chaintreau is an Assistant Professor of Computer Science at Columbia University. His research, by experience in industry, is centered on real world impact and emerging computing trends, while his training, in mathematics and theoretical computer science, is focused on guiding principles. He designed and proved the first reliable, scalable and network-fair multicast architecture while working at IBM during his Ph.D. He conducted the first measurement experience of human mobility as a communication transport tool while working for Intel and, as member of the Technical Staff of Technicolor (formerly, Thomson), showed that opportunistic caching in mobile networks can optimally take advantage of social properties.He is now working on internetworking social network services through distributed algorithms and opportunistic architecture, to vastly expand how your data and the web deal with everyday objects and your social environment. An ex student of the Ecole Normale Superieure in Paris, he earned a Ph.D. in mathematics and computer science in 2006. He has been an active member of the networking research community, serving in the program committee of ACM SIGCOMM, ACM CoNEXT, ACM SIGMETRICS, ACM MobiCom, ACM MobiHoc, ACM IMC, IEEE Infocom. He is also an editor for IEEE TMC, ACM SIGCOMM CCR, ACM SIGMOBILE MC2R.

Roxana Geambasu is an Assistant Professor of Computer Science at Columbia University. She joined Columbia in Fall 2011 after finishing her Ph.D. at the University of Washington. For her work in cloud and mobile data privacy, she received a Microsoft Research Faculty Fellowship, a ``Brillint 10" Popular Science nomination, an NSF CAREER award (all in 2014); an Honorable Mention for the inaugural Dennis M. Ritchie Doctoral Dissertation Award in 2013, a William Chan Dissertation Award in 2012, two best paper awards at top systems conferences (2009 and 2011), and the first Google Ph.D. Fellowship in Cloud Computing (2009).

LANCE D. WEILER
CO-FOUNDER OF COLUMBIA DIGITAL STORYTELLING LAB AND DIRECTOR OF EXPERIENTIAL LEARNING AND APPLIED CREATIVITY

"BALANCING STORY AND CODE"

Considered to be one of the leading voices in 21st Century storytelling, Lance Weiler has been involved in some of the most innovative projects in the space (Pandemic 1.0, Bear71, Collapsus, Lyka's Adventure and Wish for the Future).

An alumni of the Sundance Screenwriting Lab, he is recognized as a pioneer because of the way he mixes storytelling and technology. WIRED magazine named him "one of 25 people helping to re-invent entertainment and change the face of Hollywood." Always interested in experimenting with new ways to tell stories and engage audiences, Lance has designed experiences that have reached millions of people via theaters, mobile devices and online. In recognition of these storytelling innovations, BUSINESSWEEK named Lance "One of the 18 Who Changed Hollywood."

Lance sits on two World Economic Forum steering committees; one focused on the Future of Content Creation and the other examines the role of Digital Media in Shaping Culture and Governance. Lance is the Director of Experiential Learning and Applied Creativity at Columbia University as well as a founding member and Director of the Digital Storytelling Lab. He teaches a course on the art, craft and business of storytelling in the 21st Century and is currently working on a trilogy of participatory projects that centers on digital literacy and cross-generational learning.

His recent collaboration with David Cronenberg, TIFF and the CFC Media

Lab entitled Body/Mind/Change, is traveling the world for the next four years. Additionally, Lance works with large brands, agencies, studios, publishers, and gaming companies to help them shape their media holdings for the 21st Century. "Building Storyworlds: the art, craft & biz of storytelling in 21c" is a new book based on a course that Lance teaches at Columbia University on the future of storytelling.

SUSAN E. MCGREGOR
ASSISTANT PROFESSOR OF JOURNALISM

"COMPUTATION AND CROWDSOURCING FOR INVESTIGATIVE JOURNALISM"

Susan E. McGregor was the Senior Programmer on the News Graphics team at the Wall Street Journal Online for four years before joining Columbia Journalism School as an Assistant Professor in 2011, where she received an appointment as Assistant Director of the Tow Center for Digital Journalism in 2013.

While at WSJ.com, McGregor developed hundreds of interactive graphics and data applications, and was named a 2010 Gerald Loeb Award winner for her work on the the Journal's "What They Know" series. Since arriving at Columbia, she has continued to work on issues of data, graphics and digital security for journalism. In 2012 she received a Magic Grant from the Brown Institute for Media Innovation for her work on Dispatch, a mobile app for secure source communication, and in 2013 she was awarded a Knight Prototype grant to develop DataDocs, an platform for creating interactive, evergreen web videos. She holds a master's degree in Educational Communication and Technology from NYU and a bachelor's degree in Interactive Information Design from Harvard University.

ANDREW MCLAUGHLIN
PARTNER, BETWORKS | CEO, DIGG | CEO, INSTAPAPER | SENIOR FELLOW, SCHOOL OF INTERNATIONAL AND PUBLIC AFFAIRS, COLUMBIA UNIVERSITY

Andrew McLaughlin is a partner at betaworks, a technology and media start-up studio based in New York City. He also serves as CEO of Digg and Instapaper.

Andrew is chairman of the board of Access, a member of the boards of directors of Chartbeat, the Sunlight Foundation, and Public Knowledge, and a Future Tense Fellow at the New America Foundation. From 2011-2013, Andrew served on the board of Code for America.

From 2009-2011, Andrew McLaughlin was a member of President Obama's senior White House staff, serving as Deputy Chief Technology Officer of the United States. In that role, Andrew was responsible for advising the President on Internet, technology, and innovation policy, including open government, cybersecurity, online privacy and free speech, spectrum policy, federal R&D priorities, entrepreneurship, and the creation of open technology standards and platforms for health care, energy efficiency, and education. In 2008-2009, he served on the Obama/Biden presidential transition team, as a member of the Technology, Innovation and Government Reform cluster.

In 2011-12, Andrew was a Lecturer in Law at Stanford Law School, teaching a course on "Freedom of Speech in a Digitally Interconnected World," and a non-resident fellow at Stanford Law's Center for Internet & Society and at Princeton's Center for Information Technology Policy. In 2011-2013, Andrew was a member of the Commission on Innovation of the U.S. Broadcasting Board of Governors.

From 2011-2012, Andrew was EVP of Tumblr, responsible for the international, community, outreach, editorial, marketing, and support teams.

In 2011, Andrew served as the start-up executive director of Civic Commons, a Code for America initiative that helps governments build, share, and implement open-source technologies.

From 2004-2009, Andrew was Director of Global Public Policy at Google, leading the company's work on issues like freedom of expression and censorship, surveillance and law enforcement, privacy, copyrights and trademarks, regulation of Internet and telecommunications networks, wireless radio spectrum, national security, trade policy, patent reform, and online child protection. Andrew built and managed a 50-person worldwide team based in Brussels, London, Paris, Madrid, Milan, Berlin, Amsterdam, Stockholm, Dublin, Brasilia, Buenos Aires, Tokyo, Seoul, Beijing, Sydney, Ottawa, Washington, and San Francisco. Andrew

was a co-lead on Google's Africa strategy and operations.

From 1999-2003, Andrew helped launch and manage ICANN, the Internet's technical coordinating organization, serving as Vice President, Chief Policy Officer, and Chief Financial Officer. From 1998-2005, Andrew was a Senior Fellow at Harvard Law School's Berkman Center for Internet and Society. In 2002-2003, Andrew taught a course on digital democracy at Harvard Law School while working on Internet and telecom law reform projects in a number of developing countries, including Ghana, Mongolia, Kenya, Afghanistan, and South Africa. He was a co-founder of CIPESA, a technology policy think-tank and advocacy center based at Makerere University in Uganda. Andrew served as a member of the Board of Directors of Bridges.org, an international technology policy not-for-profit based in Cape Town.

After clerking on the U.S. Court of Appeals for the Eighth Circuit, Andrew started his career as a lawyer at Jenner & Block in Washington, D.C., where he focused on appellate and constitutional litigation. He was a member of the legal team that challenged the U.S. government's first Internet censorship law, resulting in the Supreme Court's landmark 1997 Internet free speech ruling in Reno vs. ACLU. From 1997-98, Andrew served as legal counsel in the U.S. House of Representatives.

In 2000, Time Magazine named Andrew one of its Digital Dozen. In 2001, he was named a Global Leader for Tomorrow by the World Economic Forum. He is a fellow of the Young Leaders Forum of the National Committee on US-China Relations.

Andrew holds a B.A. in history from Yale University, and a J.D. from Harvard Law School.

Growing up, Andrew attended Highland Elementary School in Sylvania, Ohio; Widsten Elementary School in Wayzata, Minnesota; Longfellow Elementary School and Ben Franklin Junior High School in Fargo, North Dakota; the American Community School in Cobham, Surrey, UK; and Neil A. Armstrong Senior High School in Plymouth, Minnesota, from which he graduated in 1987. Over various summers, he attended the Trollwood Performing Arts School in Fargo, the International Music Camp in Dunseith, North Dakota, and the Minnesota Institute for Talented Youth in St. Paul, at which point he had to start working lame summer jobs at places like the Sheraton Park Place Hotel, Dayton's Department Store, and the kitchen of the celebrated Red Lobster of Golden Valley.

KATHLEEN R. MCKEOWN
DIRECTOR, DATA SCIENCE INSTITUTE AND HENRY AND GERTRUDE ROTHSCHILD PROFESSOR OF COMPUTER SCIENCE

Kathleen R. McKeown is the Henry and Gertrude Rothschild Professor of Computer Science at Columbia University and she also serves as the Director of the Institute for Data Sciences and Engineering. She served as Department Chair from 1998-2003 and as Vice Dean for Research for the School of Engineering and Applied Science for two years. McKeown received a Ph.D. in Computer Science from the University of Pennsylvania in 1982 and has been at Columbia since then. Her research interests include text summarization, natural language generation, multi-media explanation, question-answering and multi-lingual applications.

In 1985 she received a National Science Foundation Presidential Young Investigator Award, in 1991 she received a National Science Foundation Faculty Award for Women, in 1994 she was selected as a AAAI Fellow, in 2003 she was elected as an ACM Fellow, and in 2012 she was selected as one of the Founding Fellows of the Association for Computational Linguistics. In 2010, she received the Anita Borg Women of Vision Award in Innovation for her work on text summarization. McKeown is also quite active nationally. She has served as President, Vice President, and Secretary-Treasurer of the Association of Computational Linguistics. She has also served as a board member of the Computing Research Association and as secretary of the board.

MATTHEW C. WAXMAN
LIVIU LIBRESCU PROFESSOR OF LAW

Matthew Waxman is an expert in national security law and international law, including issues related to executive power; international human rights and constitutional rights; military force and armed conflict; and terrorism. He holds a J.D. from Yale Law School and clerked for Associate Supreme Court Justice David H. Souter and Judge Joel M. Flaum of the U.S. Court of Appeals for the Seventh Circuit. Before joining the Columbia faculty, he served in senior positions at the U.S. State Department, Department of Defense and National Security Council. Professor Waxman was a Fulbright Scholar to the United Kingdom, where he studied international relations and military history. He is a member of the Council on Foreign Relations, where he also serves as Adjunct Senior Fellow for Law & Foreign Policy, and he is the Co-Chair of the Cybersecurity Center at the Columbia Data Science Institute. matthew c. waxman liviu librescu professor of law

SHARON K. SPUTZ
DIRECTOR OF STRATEGIC MARKETING, DATA SCIENCE INSTITUTE

Sharon has combined experience in business strategy and technical research. Prior to joining the Institute in 2014, Sharon spent 11 years at BAE Systems identifying and pursuing new business opportunities in leading edge technology. She began her career at Bell Laboratories managing and coordinating material characterization as well as developing new and innovative tools. Sharon moved on to become the Strategic Marketing Manager at Lucent/Agere with a variety of responsibilities: from leading the technical evaluation of merger and acquisition candidates to defining product planning for the optical networking group and components. Sharon received a Bachelor of Science in Physics from State University of NY at Binghamton and a Masters of Science in Physics from Stevens Institute of Technology.

PARTNERING WITH THE CAT

CENTER for ADVANCED INFORMATION MANAGEMENT
a New York State Center for Advanced Technology at Columbia University

What is the Center for Advanced Information Management?

- The Center for Advanced Technology (CAT) at Columbia University is known as CAIM. The CAT program is funded by New York State (NYS) to promote collaboration between universities and NYS companies. Our goal is to create a direct tangible economic impact/benefit, for these companies through a partnership with Columbia University.
- The CAT program is funded by NYSTAR, a division of NYS Empire State Development.

What are the benefits of partnering with the CAT for faculty?

- Matching grants for company sponsored projects at Columbia
- Support for locating corporate sponsors for research and development projects
- Entrepreneurship activities and events (ie. HealthTech Assembly, AWS Grant Program)
- Joint preparation of proposals to third party sources (Examples: SBIR/STTR)
- Consulting arrangements with companies
- Education/training programs and events

CAT funds are available to help support and augment budgets provided by the company. The requirements for a CAT grant are (1) a joint project with a focus on science/technology; (2) a reasonable project time frame and budget estimate; (3) the promise of a successful commercial product; and (4) the cooperation of company senior management to provide documentation of benefits to the company for semi-annual reports

Matching Grant Application Process

- The application for a company co-supported grant is straightforward (2 page form + budget) and reviewed internally at Columbia University.
- Decisions on CAT grant applications are made quickly (usually 1-2 weeks)
- Funds are made available for the faculty member in a project account in their department.
- Grant awards typical are for $25-50k/year but may be higher depending on the projected economic impact to the company.
- CAT grants are charged a special low 15% (on salary + fringe benefits) indirect cost rate, about ¼ of the standard federal rate.
- Company matching funds may also be eligible for the special low 15% IDC rate.

For more information on our various programs, please see:

www.cat.columbia.edu
@ColumbiaUCAT
630 West 168th Street, BB-130 New York, NY 10032
(212) 305-2944

THE DATA SCIENCE INSTITUTE WOULD LIKE TO ACKNOWLEDGE THE FOLLOWING ORGANIZATIONS FOR THEIR SUPPORT AND ENGAGEMENT WHICH HAVE MADE THE DATA ON A MISSION EVENTS POSSIBLE:

DATA SCIENCE INDUSTRY AFFILIATES

Bloomberg

Booz | Allen | Hamilton

Fitek LLC

GE

Goldman Sachs

mediaocean

synergic partners

VSP Global

FRIENDS OF THE INSTITUTE

Microsoft Research

YAHOO!

GORDON AND BETTY MOORE FOUNDATION

ALFRED P. SLOAN FOUNDATION

COLUMBIA UNIVERSITY
ROADS PROVOST IGNITION FUNDING

Columbia University
Data Science Institute

Calendar of institute events

COLLOQUIUM SERIES
I³ INNOVATION SEMINARS
SPEED NETWORKING
TECH TALKS AND HACKATHONS
datascience.columbia.edu/events/calendar

Industry Engagement

INDUSTRY AFFILIATES PROGRAM
CAPSTONE PROJECT
MENTORING AND RECRUITING
SPONSORED RESEARCH
industry.idse.columbia.edu

Data Science Education

MASTER OF SCIENCE
CERTIFICATE OF PROFESSIONAL ACHIEVEMENT
datascience.columbia.edu/academics-idse
datascience.columbia.edu/apply

DATASCIENCE.COLUMBIA.EDU | DATASCIENCE@COLUMBIA.EDU | (212) 854-5660