Data Science Research Constellation

Key to
Knowledge
Discovery

January 2021

Table of Contents

Data Science Research Constellation (DSRC)

Preamble

The emerging field of *data science*, a synthesis of theories and methodologies from multiple disciplines, is having a transformative impact reaching from basic science to the tasks of everyday life. It stands nearly unique among academic endeavors in both spurring innovative research discovery while driving daily decision-making in a wide-range of relevant practical domains. Its beneficiaries range from government and industrial organizations to individual citizens. Data science finds itself at the heart of cyber-security, environmental monitoring, bio-medical research, and business creation and operations. Moreover, it is playing an increasingly important role in knowledge discovery across a multitude of scientific disciplines. Succinctly put, harnessing the power of data science will be key to countless successful endeavors within the governmental, academic, commercial, and consumer sectors. Recognition of the transformative role of data science is reflected in the National Science Foundation (NSF) designating *Harnessing the Data Revolution* as one of its 10 Big Ideas[1].

To be more specific, data science is an amalgamation of theories and methodologies from core disciplines such as computer science, mathematics, and statistics, with its tool-set enriched and informed through contributions from application areas such as bioinformatics, business analytics, engineering, and medicine. Its impact is not limited to science and engineering, but has spread to language arts as well as other areas of humanities. In many academic institutions, data science is practiced piecemeal, within the silos of its core component disciplines and application areas. A report by the National Academies of Science, Engineering, and Medicine[2] points to the complementary nature of the strengths and weaknesses of the individual components of data science, with the implication that the true power of data science comes through the integration of its constituent parts.

In addition, there is growing concern about the pitfalls of using data science tools without thoughtful consideration and safeguards, leading to enhancing of existing biases or adding new ones, ethical violations, and negative societal impacts. Thus, the focus of data science research and education should not be merely the creation of data science tools, but also on their ethical implementation and the study of its societal impacts and measures to mitigate such impacts. Clearly, data science should be viewed holistically, with its foundational components complemented with disciplines that help address these concerns. Thus, it is the intention of this research group to cut across a broad swath of multiple disciplines.

Many leading academic institutions have recognized the importance of creating an environment that brings together researchers in multiple disciplines, with such coalescing acting as a catalyst for the precipitation of new scientific breakthroughs and technologies. This idea that an

[1] https://w ww.nsf.gov/news/special_reports/big_ideas/

[2] National Academies of Sciences, Engineering, and Medicine. *Envisioning The Data Science Discipline: The Undergraduate Perspective: Interim Report.* Washington, DC: : National Academies Press, 20

environment which brings together researchers in complementary areas will act as an incubator for new ideas and discoveries forms the motivation for forming this research constellation.

Vision

The vision of the data science research constellation is to create an interconnected network of researchers engaged in both foundational and applied areas of data science and to promote innovative research that transcend disciplinary boundaries, thereby creating an agile community of Missouri S&T researchers enabled to successfully respond to opportunities and challenges not only in the data science arena, but also in scientific, engineering, and business endeavors that benefit from it.

Major Goals

- Establish a collaborative research network that enables its members to engage in the exchange and synthesis of ideas and methods, thereby creating an incubator for transformational research
- Bring together researchers in application and foundational areas of data science
- Create a network of researchers who can respond quickly and effectively to large-scale funding opportunities and challenges
- Within the framework of the big umbrella of data science, create subgroups that work on synergistic areas such as foundational research in data science, high-performance computing, data and society, financial and business analytics, intelligence systems etc.
- Establish and strengthen connections to research center, signature research areas, and the newly proposed Kummer Institute research centers
- Create resources and opportunities for both undergraduate and graduate students to engage in data science research that will prepare them for the modern workplace or for future research careers
- Establish cross-disciplinary data science emphasis areas in undergraduate and graduate degree programs, including those at the Kummer School of Entrepreneurship, which will integrate data science research with the university's core mission of education.

Short-term Action Items

- Establish a website listing data scientists in our group with their expertise and research projects. *This will enable members to find collaborators for existing projects and writing grant proposals.*
- Establish a seminar series for short presentations of members' work. *This will enable members to get to know about each other's expertise as well as research projects and learn about new research techniques and methods. The talks will be recorded and archived.*

- Create opportunities for informal one-on-one as well as group discussion in a social setting. *This would facilitate an environment for networking and exchange of ideas.*
- Explore opportunities for large-scale funded projects. *This will enable our researchers to respond quickly to federal agency proposal calls for such projects.*
- Explore funding opportunities for undergraduate research.
- Engage graduate and undergraduate students in data science research. *For example, organize workshops, short training sessions, certificate programs, and similar programs.*
- Establish a colloquium series where national and international leaders in both foundational and application areas of data science are invited to give talks.
- Connect with industry partners to seek out potential funding opportunities well as internship opportunities for students. *These connections can also lead to novel research problems and long-term partnerships.*

Long-Term Action Items

- Establish a Center (or Institute) for Data Science Research
- Establish a *Data Science Commons*, a convening space and other resources to enable cross-exchange of expertise and encourage the emergence of new research ideas and collaborations
- Establish collaborative partnerships with other Data Science Centers in the U.S. and abroad
- Establish scholar exchanges with select institutions.

Current Strengths

As the Venn diagram in Figure 1 shows, Data Science is not a discipline that fits into any existing department on the Missouri S&T campus. This is also the case with most universities across the Nation. Each member of the data science research group belongs to one of the sets in the Venn diagram, but may be doing work across these disciplines. This is a significant strength of this group. Moreover, many in our group belong to existing research centers.

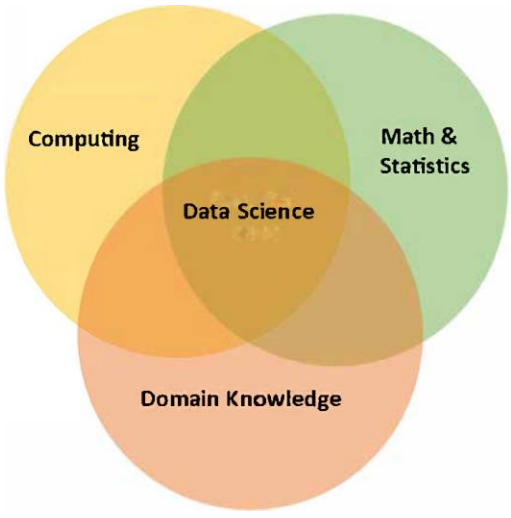

Figure 1 Main Components of Data Science

There are multiple groups on campus conducting research on fundamental areas of data science, on data science applications, as well as its ethical implications and impact on society, yet there is only a limited exchange of research ideas and technologies across groups. The goals and action items proposed herein are aimed at bringing these groups together to facilitate convergence between their technologies and create a more vibrant data science research community to the benefit of all researchers, emulating what was done in three leading institutions (see *Creating Institutional Change in Data Science*[3]), thus positioning Missouri S&T as a leader in data science.

[3] file:///C:/Users/vsam/AppData/Local/Temp/Creating_Institutional_Change.pdf

Constellation Membership and Research Interests

The current membership of the DSRC illustrate its strength. Table 1 given below lists the members of the Data Science Constellation together with their research interests and contact information. They are categorized into the three main components shown in the Venn diagram above using the same color code. Research Center affiliations are given on the last column.

Table 1. Current Members of the Data Science Research Constellation

Name	Department	Research Interests	Contact/Center Membership
Akim Adekpedjou	Mathematics and Statistics	Recurrent Event data Analysis, stochastics processes, survival analysis	akima@mst.edu MICAMS
Burns, Devin	Psychological Science	Decision making, Individual differences, Augmented Perception	burnsde@mst.edu (573) 341-4807 ISC, CSTS
Cavaglia, Marco	Physics	Astrophysics, gravitational physics, data analysis for large-scale experiments, machine learning, scientific collaboration	cavagliam@mst.edu (573) 341-4781 ISC, S&T IMAC, LIGO
Chen, Langtao	Business and Information Technology	Business analytics, social media, online communities, user-generated content, health informatics, cybersecurity, big data, machine learning	chenla@mst.edu (573) 341-4418
Das, Sajal	Computer science	Cyber-Physical Systems, Security and Privacy, Smart Environments (Smart City, Energy, Healthcare), IoTs, Wireless Sensor Networks, Mobile and Pervasive Computing, Big Data Analytics, Parallel and Cloud Computing, Social Networks, Systems Biology, Graph Theory and Game Theory	sdas@mst.edu (573) 341-7708 ISC
Emdadi, Arezoo	Materials Science and Engineering	Computational mechanics/material science including phase-field modeling (crack propagation, solidification, oxidation, material microstructure), phase-field crystal modeling, finite element analysis, fracture mechanics, and ICME.	emdadia@mst.edu (573)341-4796 MRC
Fikru, Mahelet	Economics	Energy, photovoltaic adoption, electricity bill savings, cost-benefit analysis, policy, mergers and acquisitions, regression analysis	fikruma@mst.edu (573) 341-6495 CREE
Gao, Stephen S.	Geosciences and Geological and Petroleum Engineering	Signal detection, data mining, time series analysis, seismology, earthquakes, computer-aided tomography, machine learning	sgao@mst.edu 573-341-6676 HPCC

Table 1. Current Members of the Data Science Research Constellation (continued)

Name	Department	Research Interests	Contact/Center Membership
He, Xiaoming	Mathematics and Statistics	Data-enabled modeling, data assimilation, interface problems, numerical methods for partial differential equations, computational fluid dynamics, computational plasma physics	hex@mst.edu (573) 408-0417 HPCC, MICAMS
Hu, Wenqing	Mathematics and Statistics	Differential and Difference Equations, Statistical Learning, Data Science	huwen@mst.edu (573) 341-4650
Hilgers, Michael	Business and Information Technology	Information visualization, modeling and simulation applied to business analytics and data science, quantitative finance theory and application, virtual reality research for first responder training	hilgers@mst.edu (573) 201-1710
Liu, Kelley	Geosciences and Geological and Petroleum Engineering	Digital signal processing and analysis, geophysical modeling and inversion, subsurface imaging	liukh@mst.edu (573) 341-6724 HPCC
Luo, Tony	Computer Science	Internet of Things, Machine Learning, Security, Trust, and Privacy	tluo@mst.edu (573) 341-4788
Madria, Sanjay	Computer Science	Cloud Computing, Security, Wireless Computing and Mobile Data Management, Data Analytics, Big data	madrias@mst.edu 5733414856 ISC
Maurer, Jeremy	Geological Sciences and Geological and Petroleum Engineering	Remote sensing, uncertainty quantification, big data, machine learning, optimization	jmaurer@mst.edu (573) 341-4852
Mormile, Melanie	Biological Sciences	Environmental microbiology, extremophiles, halophiles, biotechnology	mmormile@mst.edu (573) 341-6346 CREE
Murphy, Jason	Mathematics and Statistics	Harmonic analysis, analysis of PDEs, nonlinear wave equations, inverse problems, medical imaging, nonlinear optics	jason.murphy@mst.edu (573) 341-4654
Nadendla, Sid	Computer Science	Cyber-Physical-Human Systems; Statistical Inference & Machine Learning; Trust and Influence; Algorithmic Game Theory; Security, Fairness, Transparency	nadendla@mst.edu (573) 341-4090 ISC, CII
Nah, Fiona	Business and Information Technology	Human-computer interaction, user experience, Neuro-IS (eye-tracking & EEG), usable privacy and security, gamification, virtual communities and collaboration, digital commerce, meta-analysis	nahf@mst.edu (573)341-6996 ISC, CSTS
Obafemi-Ajayi, Tayo	Electrical and Computer Engineering	Machine learning; intelligent systems; data analytics; bioinformatics.	towd2@mst.edu (417) 837-2317
Olbricht, Gayla	Mathematics and Statistics	Statistics/data analytics (specializing in biological/biomed applications); Machine learning; bioinformatics	olbrichtg@mst.edu (573) 341-4913 ISC, HPCC, CSCMBC

Table 1. Current Members of the Data Science Research Constellation (continued)

Name	Department	Research Interests	Contact/Center Membership
Saito, Shun	Physics	Cosmology, Dark Energy, Galaxy evolution, Data compression, Statistical Inference	saitos@mst.edu (573) 201-3601
Samaranayake, V.A. (Sam)	Mathematics and Statistics	Time Series, Prediction Models, Reliability, Statistical/data analytic applications in biology, social sciences, and engineering	vsam@mst.edu 573-341-4658 MICAMS, CBR,CSTS, HPCC, CSCMBC
Sarangapani, Jagannathan	Electrical and Computer engineering	Systems and control; neural network control; event-triggered control/cyber-physical systems; resilience/prognostics; autonomous systems/robotics.	sarangap@mst.edu (573) 341-6775 ISC
Shank, Daniel B.	Psychology	Social Psychology, Technology, Human-Computer Interaction, Artificial Agents, Morality, Social Interactions	shankd@mst.edu (573) 341-4823 CSTS, ISC
Shen, Ting	Psychological Science	Quantitative research methods, large-scale assessment data, psychometrics, empirical research in psychology and education	tingshen@mst.edu (573)341-4696 CSTS
Siau, Keng	Business and Information Technology	Data Science, Sentiment Analytics, Emotion Analytics	siauk@mst.edu (573) 341-7262 ISC
Singler, John	Mathematics and Statistics	Image and signal processing; computational intelligence; data fusion; automation; bioinformatics.	singlerj@mst.edu (573) 341-4648 HPCC, MICAMS, CSMBC
Stanley, Joe	Electrical and Computer Engineering	Image and signal processing; computational intelligence; deep learning; data fusion; automation; bioinformatics.	stanleyj@mst.edu (573) 341-6896
Thimgan, Matthew	Biological Science	Sleep, Cognitive performance, Biomarkers, Health metrics	thimgan@mst.edu (573) 341-7190 CBR, CSCMBC
Tripathy, Ardhendu	Computer Science	Multi-armed bandits, sequential analysis, statistical learning theory, algorithms	astripathy@mst.edu (515) 509-0388
We, Xuerong (Meggie)	Mathematics and Statistics	Nonlinear and nonparametric regression, regression graphics, computational statistics and statistical genetics, with an emphasis on sufficient dimension reduction in the context of regression	wenx@mst.edu (573) 341-6209 MICAMS

Table 1. Current Members of the Data Science Research Constellation (continued)

Name	Department	Research Interests	Contact/Center Membership
Wunsch, Donald	Electrical and Computer Engineering	Clustering; neural networks; reinforcement learning; approximate dynamic programming; adaptive dynamic programming; lifelong learning; explainability and ethics of AI..	dwunsch@mst.edu (573) 341-4521 ISC, CBR, HPCC, CIES, CSCMBC
Zhang, Yanzhi	Mathematics and Statistics	Data-driven modeling, Machine learning algorithms and applications, Optimal control Numerical modeling and simulations with partial differential equations	zhangyanz@mst.edu (573)-341-4651 HPCC, MICAMS

*Center Abbreviations are given in Table 2

Table 2. Research Center abbreviations

CBR - Center for Biomedical Research	EMRCGC - Energetic Materials, Rock Characterization, and Geomechanical Center
HPCC - Center for High Performance Computing	ISC - Intelligent Systems Center
CIES - Center for Infrastructure Engineering Studies	MRC - Materials Research Center
CII - Center for intelligent Infrastructure	CSCMBC - Center for Statistical and Computational Modeling of Biological Complexity
CREE - Center for Research in Energy and Environment	MICAMS - Missouri Institute of Computational and Applied Mathematical Sciences

Key:	Class A centers	Class C centers

Current and Potential Future Collaborative Connections to Missouri S&T Research Centers, Consortia, and the Proposed Kummer Institute Centers.

A significant proportion of the Data Science Research Constellation (DSRC) members are active in other Missouri S&T research centers. This connection to existing research endeavors is another strength of the DSRC. More importantly, since data science is becoming an important tool for scientific knowledge discovery and industrial development, DSRC can play a critical role in the success of the Missouri S&T Research Centers, the proposed Kummer Institute Centers, as well as some of the S&T Research Consortia. Figure 2 highlights the existing as well as potential for further research collaborations between DSRC and these centers. The strong research focused benefit the campus and KI research centers gain from individual DSRC members can be further enhanced by the formation of the DSRC that will act as a catalyst for the creation of new scientific breakthroughs and technologies through transdisciplinary research.

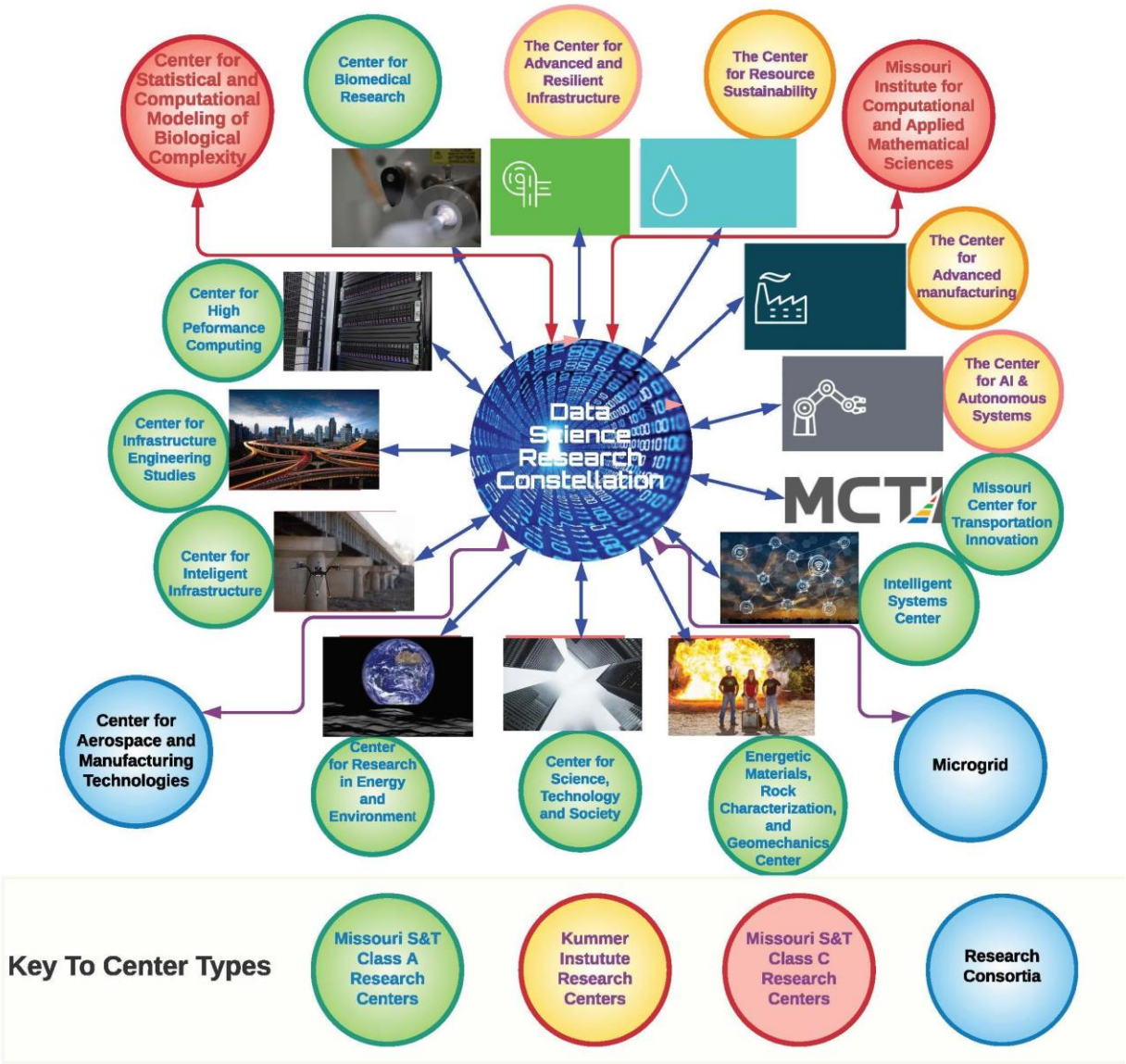

Figure 2. Current and Potential Future Collaborative Connections to Missouri S&T Research Centers, Consortia, and the Proposed Kummer Institute Centers.

Tables 3 and 4 given below highlight the nature of the collaborative research that is possible between the current DSRC members and those connected to the S&T and Kummer Institute (KI) research centers. While there is ongoing research between some DSRC members and the Missouri S&T centers and consortia, an existing strength, the formation of the DSRC as an active incubator of new research ideas will further enhance and add to these synergistic links. Moreover, the DSRC will be able to play a key role in the proposed KI centers. In addition, these collaborations can generate new research ideas and funding opportunities.

Table 3. Potential Research Links between DSRC and the Missouri S&T Research Centers and Consortia

S&T Research Centers	Potential Collaborative Research
Center for Biomedical Research (CBR)	Property-composition analysis of biomaterials using Ai-based/ statistical models. Predictive modeling. Data driven modeling of biomaterials. Statistical modeling of patient outcomes. Examples: https://pubs.rsc.org/en/journals/articlecollectionlanding?sercode=me&themeid=aafdd74e-6b01-43ec-8989-b79c34a1ef43.
Center for High Performance Computing (HPCC)	The high performance computer facilities would be essential to some of the data science research. Collaborative research with members of the HPCC will include developing efficient and fast algorithms to facilitate computationally intensive research.
Center for Infrastructure Engineering Studies (CIES)	Analysis of censor data for infrastructure monitoring. Predictive modeling of material performance. Examples: https://www.stevens.edu/news/data-science-and-machine-learning-drive-innovation-civil-engineering-stevens.
Center for intelligent Infrastructure (CII)	Integrating automation, informatics, and actuation into the life-cycle assessment, performance prediction, risk analysis, and post-disaster recovery of infrastructure systems through multi-scale modeling and sensing.
Center for Research in Energy and Environment (CREE)	Environmental monitoring generate large quantities of spatio-temporal data requiring innovative data analytic methods for extracting meaningful information. Real-time monitoring requires AI-based methods and statistical methods to isolate meaningful signals from noisy data. Energy efficiencies can be optimized through data science techniques that model usage patterns and well fluctuations in the energy demands. Examples: https://www.discoverdatascience.org/industries/clean-energy/.
Center for Science, Technology, and Society (CSTS)	The way technology affects society and vice-versa is a topic that can be best understood by analyzing the vast amount of data that can be extracted from social media as well as smart devices. Data science is already playing a major role in these investigations. Another important aspect is the unintended consequences of the use of data science tools such as racial biases and ethical issues that arise due to unmodulated use of them. Research projects that study these impacts will bring together CSTS members and those in the DSRC.

Table 3. Potential Research Links between DSRC and the S&T Research Centers and Consortia (continued)

S&T Research Centers	Potential Collaborative Research
Energetic Materials, Rock Characterization, and Geomechanical Center (EMRCGC)	Screening energetic materials using AI and statistical methods. Mining large databases to extract new knowledge to facilitate the development of novel energetic materials, smart monitoring of the physiological impact of explosives, collaborative work in 4D data integration, modeling of geospatial data with a temporal component, computational and data driven geomechanics studies.
Intelligent Systems Center (ISC)	Joint development of novel AI tools, adaptation of data science tools for system monitoring and control, joint research involving translational research in data science, collaborative work in developing smart devices, smart structures, and smart cities.
Materials Research Center (MRC)	Data driven materials science, extraction of new knowledge from large materials data sets, property-composition and structure-property modeling, simulation, integration of experimental and computational data, materials testing and degradation studies, data analytic and simulation-based approaches to optimizing material performance, providing statistical framework for designing experiments for testing biomaterials.
Missouri Center for Transportation Innovation (MCTI)	Data-driven approaches to optimizing transportation systems, building simulation models, developing real-time monitoring systems for transportation infrastructure and the analysis of transportation data.
Center for Aerospace Manufacturing Technologies	Developing robust automated control systems, simulation modeling of individual parts as well as complex systems, reliability modeling using statistical and machine learning tools, use of computer vision for automated quality inspection, automated process control and fault detection using machine learning algorithms.
Microgrid	Analyzing data streams from the real-time monitoring of power management building and solar homes in the ecovillage to understand underlying usage patterns, their relationship to human activity and weather; building simulation models to understand how this complex system behaves under stress.

Table 4. Potential Research Links between DSRC and the Kummer Institute Centers

KI Research Center	Potential Collaborative Research
Center for Advanced Resilient Infrastructure	Developing advanced monitoring systems for business infrastructure, enhancing the use of business informatics and decision making tools, developing data-informed strategies for enhancing the resilience of business infrastructure.
The Center for Resource Sustainability	Developing data-driven approaches to efficient extraction of mineral and fuel resources while reducing environmental impact and optimizing resource use and allocation.
Center for advanced Manufacturing	Data driven approaches to optimizing manufacturing, integration of experimental and computational data, developing robust control systems based on machine learning algorithms, automated additive manufacturing, using cloud-based software to automate manufacturing across multiple machines and plants, use of computer vision for automated quality inspection and fault detection.
Center for Artificial Intelligence and Autonomous Systems	The DSRC members can play a major role in research conducted by this center. The members of the DSRC who are conducting foundational research in the two theoretical components of data science can develop new methodologies and algorithms to address emerging challenges and the domain specific scientists and engineers within DSRC can directly work on projects undertaken by this center.

Potential Funding Opportunities

There are a considerable number of funding opportunities that are directly related to data science research. In addition, data science techniques are becoming increasingly important in engineering and scientific research and thus, grant opportunities in these areas that require a data science component have become common.

The funding opportunities for basic research in data science are mainly focused within the National Science Foundation initiatives (e.g. Harnessing the Data Revolution), but opportunities for application areas of data science extend across all national funding agencies. These range from the joint NSF-NIH initiative: *Smart Health and biomedical Research in the Era of Artificial Intelligence and Advanced Data Science* to Navy Research Lab's Broad Agency Announcements related to Big Data and high performance computing. In addition, data intensive research initiatives are not limited to what is considered data science in a narrow sense. For example, National Science Foundation's Division of Mathematical Sciences (NSF/DMS) is partnering with the National Institute of General Medical Sciences (NIGMS)to promote research at the interface of the biological, biomedical, and mathematical sciences, to exploit the extraordinary growth of available data.

Given below is a sample of the funding opportunities that are available to the DSRC members.

1. National Science Foundation Initiatives

Here are select NSF funding opportunities (posted in 2020):

- Harnessing the Data Revolution (HDR): Data Science Corps (DSC)
- Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering
- Methodology, Measurement, and Statistics
- Computational and Data-Enabled Science and Engineering
- Geoinformatics (Cyberinfrastructure for easy access to high-quality data, visualization tools, and modeling and analysis codes …)
- Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science (see NIH listing below)
- Human Networks and Data Science
- Dear Colleague Letter: Data Science Activities for the Civil, Mechanical and Manufacturing Innovation Communities (supplement to existing grants)
- Algorithms for Threat Detection (ATD)
- Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science
- Leveraging Big Data Science to Elucidate the Mechanisms of HIV Activity and Interaction with Substance Use Disorder

2. National Institute of Health and related Agencies

Here are a few examples of NIH funding opportunities directly linked to data science. There are, however, numerous NIH funding opportunities where data science tools can help in the scientific discovery process.

- Data Science Research: Personal Health Libraries for Consumers and Patients
- Smart Health and biomedical Research in the Era of Artificial Intelligence and Advanced Data Science (NIH-NSF interagency initiative)
- Notice of special interest: The Application of Big Data Analytics to Drug Abuse Research (can propose the analysis of existing data)
- Notice of special interest: Modeling Social Contagion of Substance Use Epidemics (Applications should make use of large data sets and data science approaches to develop computational models of social networks ...)
- Joint DMS/NIGMS Initiative to Support Research at the Interface of the Biological and Mathematical Sciences (DMS/NIGMS)
- Summer Institute for Research Education in Biostatistics and Data Science (R25)

3. Department of Defense and Department of Energy

- NRL Broad Agency Announcement (proposals related to Big Data and one of Information management and decision architecture, mathematical foundations of high assurance computing, distributed computing/network architecture, high performance computing, atmospheric effects, …)
- Notice of Intent to Issue a Funding Opportunity for the High Performance Computing for Energy Innovation Program
- National Library of Medicine (NLM) Research Grants in Biomedical Informatics and Data Science
- Investigator Initiated Research in Computational Genomics and Data Science
- Scientific Discovery Through Advanced Computing: Partnerships in Basic Energy Sciences

Ongoing Funded Data Science Research Projects of Members

Given below are the currently active funded research projects of these select constellation members.

Marco Cavaglia

- Improving Data Quality of Advanced LIGO Gravitational-Wave Searches. NSF award Number: PHY-1921006; Principal Investigator: Marco Cavaglia. Start Date: 01/01/2019. Award Amount: $348,181.00, (100%).
- WoU-MMA: Enabling Multi-Messenger Astrophysics with Advanced LIGO: from Detector Calibration to Interpretation of Gravitational-Wave SIgnals. NSF award Number: PHY-2011334; Principal Investigator:Marco Cavaglia. Start Date: 08/01/2020. Award Amount: $79,079.00, (33%).

Xiaoming He

- National Science Foundation, DMS-1722647, “Collaborative Research: Data-enabled modeling, numerical method, and data assimilation for coupling dual porosity flow with free flow”, August 2017 - July 2021, Xiaoming He (PI), Baojun Bai (Co-PI), and Mingzhen Wei (Co-PI), total amount: $180,000, (70%).
- National Science Foundation, DMS-1818642, “Collaborative Research: Models, algorithms, simulations, and applications for two-phase ferrofluid flows in contact with a solid surface”, August 2018 - July 2021, Xiaoming He (PI) and Cheng Wang (Co-PI), total amount: $160,000, (70%).

Mahelet Fikru

- Sloan Foundation, “Substitutes versus complements: effect of renewable procurement on installation decisions”, Casey Canfield (PI), Mahelet Fikru (Co-PI), August 2020 - December 2022. Total amount: $150,000, (100%).

Daniel Shank

- CI Canfield (PI), CH Dagli, DB Shank, MA Schnitzler, K Lentine, H Randall. Oct 2020 – Sept 2021. Teaming Transplant Professionals and Artificial Intelligence Tools to Reduce Kidney Discard. *National Science Foundation.* $150,000. Shared credit: $37,500 (25%).
- DB Shank (PI). Apr 2019 – Mar 2022. Mind Perception and Morality of Artificial Intelligence in Social Interaction. *Army Research Office.* $229,681 (100%).
- DB Shank (PI). Apr 2019 – Mar 2021. Modeling the Perceptions of Teams Based on Team Member Behavior: Human versus Autonomous Agent Team Members. *Leonard Wood Institute.* $72,566 (100%).

Sanjay Madria

- Sole PI, BDD: Efficient and Scalable Collection, Analytics and Processing of Big Data for Disaster Applications, NSF, $347K, 4/2015-9/2020.
- Sole PI, Machine Learning for Secure & Resilient Information Management in Combat Cloud, AFRL, 2020-2021, $500K.
- PI, A Machine Learning based ICME Framework for Optimizing Metal Additive Manufacturing Process Parameters to Enhance Fatigue Performance, $20K, BIC Project, 7/2020-6/2021.
- PI, Big Data and Machine Learning for Security and Safety, DoE, 10/2018-9/2022, $650K.

Areas for Improvement

While Missouri S&T has a core group of researchers working in both foundational and application areas of data science, with research innovations and funding successes associated with their work, there are areas for improvement. These are summarized below.

- In order to be highly competitive when seeking multi-million dollar grants for establishing data science centers/institutes and similar projects, strengthening the core group of data scientists with strategic hires is desirable
- As evidenced from schools that have established data science research centers, their success as research incubators is linked to having a physical presence where participants from multiple disciplines can gather and interact. Such a physical space needs to be established
- Successful research goes hand-in-hand with having a vibrant graduate program and thus, interdisciplinary graduate programs in data science should be established
- Resources will also be needed to bring in nationally known data scientists as visitors, both on a short and long term basis.

Appendix

An Overview of Member Research Interests

Name: Akim Adekpedjou

Department: Mathematics & Statistics

Key words: Spatial statistics, Recurrent failure, Rank-based inference, Longevity risk pricing, Ruin theory, and Targeted learning in data science

A brief overview of research: My main research area is developing dynamic statistical models for the analysis of single and recurrent events time. In the last few years I have focused on other topics such as spatial statistics, survey sampling, and actuarial science (specifically ruin related problems and longevity risk pricing). Since all these topics deal with large datasets, I am currently investigating, in various projects with my PhD students, ways to use data science tools for a better understanding of data arising from these areas. Rank-based inference, which has been shown to be robust with respect to errors, is also employed in all these methodologies in order to assess efficiency gained with respect to existing approaches.

Name: Devin Burns

Department: Psychological Science

Key words: Decision making, Individual differences, Augmented Perception

A brief overview of research: It's eclectic! Dr. Balakrishnan and I have pursued using quantum probability models to predict decision making behavior that deviates from classical norms of rationality. Dr. Song and I are working on improving human perception of balance with augmented perception devices and measuring how humans communicate information through physical force. Something this community might be interested in is a project Dr. Shen and I are conducting examining CET survey results investigating gender differences, class size effects, etc. I would like to create a dataset to validate our measuring of teaching effectiveness by using standardized test scores and grades in subsequent classes as objective criteria of effectiveness.

Name: Marco Cavaglia

Department: Physics

Key Words: Astrophysics, gravitational physics, data analysis for large-scale experiments, detector data, machine learning.

A brief description of research: My research is in experimental and theoretical gravitational physics, astrophysics, and data analysis for large-scale physics experiments including but not limited to signal processing methods and machine learning algorithms. I am a senior member of the Laser Interferometer Gravitational-wave Observatory (LIGO) Scientific Collaboration (LSC), where I served as assistant spokesperson from 2012 to 2017 and I am now serving as the co-chair of the LSC second largest data analysis group. LIGO is the largest NSF-funded experiment to date, comprising two large-scale detectors in Washington and Louisiana. The LSC is an international team of over 1300 researchers, scientists and engineers from over 100 institutions in 20 countries that exploits LIGO science data. The LSC and its European partner Virgo announced the first direct detection of cosmic gravitational waves signals in February 2016, proving Einstein's "right." My research group contributes to the science of the LSC by studying the astrophysics of black holes and other high-energetic celestial phenomena, analysis of LIGO observational data, and detector characterization and calibration with dedicated work at the LIGO sites.

Name: Langtao Chen

Department: Business and Information Technology

Key words: Business analytics, social media, online communities, user-generated content, health informatics, cybersecurity, big data, machine learning

A brief overview of research: My research focuses on business analytics, online communities, social media, user-generated content, health information technology, and cybersecurity. Methodologies such as machine learning (supervised and unsupervised), social network analysis, and econometrics are widely used in my research. Particularly, I'm interested in analyzing archival or digital trace data and extracting meaningful and interesting constructs to model/explain individual or organizational behaviors. My research has been published in journals such as Journal of Management Information Systems, Journal of the Association for Information Systems, Decision Support Systems, among others.

Name: Sajal Das

Department: Computer Science

Key Words: Cyber-Physical Systems; Security and Privacy; Smart Environments; IoTs; Wireless and Sensor Networks; Mobile and Pervasive Computing; Big Data Analytics; Parallel, Distributed, and Cloud Computing; Social Networks; Systems Biology; Applied Graph Theory and Game Theory.

A Brief overview of research: My research interests are in Cyber-physical systems, Security and privacy, Smart environments (Smart city, energy, healthcare), IoTs, Wireless sensor networks, Mobile and pervasive computing, Big data analytics, Parallel and cloud computing, Social networks, Systems biology, Graph theory and Game theory. My recent funded research include Secure and Trustworthy Framework for Integrated Energy and Mobility in Smart Connected Communities, Breakthrough: Securing Smart Grid by Understanding Communications Infrastructure Dependencies, Threat Assessment Tools for Management-Coupled Cyber- and Physical- Infrastructures, Sparsification-based Approach for Analyzing Network Dynamics, National Science Foundation (NSF) XPS: SPANDAN: Scalable Parallel Algorithms for Network Dynamics Analysis, and Crafting a Human-centric Environment to Support Human Health Needs.

Name: Arezoo Emdadi

Department: Materials Science and Engineering

Key words: Computational mechanics/material science including phase-field modeling (crack propagation, solidification, oxidation, material microstructure), phase-field crystal modeling, finite element analysis, fracture mechanics, and ICME.

A brief overview of research: My research area has centered around analytical (mathematical) and numerical modeling of material properties in multi time/length scales. Application of engineering materials for extreme environments, onset of novel processing methods, and new fabrication processes are some of my key research areas with respect to modeling material properties suited to the needs of specific applications. For example, I use phase-field model to study microstructural evolution under different process/fabrication methods like alloy additive manufacturing or plasma arc welding of ultra-high-temperature ceramics (UHTC); this can help understanding process-material-properties bridges and make a solid foundation for developing tailored materials. Machine learning, data mining and artificial intelligence are exciting tools that have entered the material science toolbox in recent years. I am interested in utilizing these powerful tools in related materials science applications such as discovery of stable materials and the prediction of their crystal structure, scaling up from the laboratory to manufacturing, and tailored parameters.

Name: Mahelet Fikru

Department: Economics

Key words: Energy Use, Energy Generation, Photovoltaics, Savings, Cost-Benefit Analysis, Electricity Sector, Environmental Policy

A brief overview of research: My research area is in energy and environmental economics as well as industrial organization. I am interested in the adoption of residential solar photovoltaics and performing a cost-benefit analysis from the individual as well as social point view. For example, I use household surveys to understand the economic and behavioral motivations to the adoption of solar photovoltaics. I use high-frequency load and solar generation data to measure savings attributable to solar photovoltaics and examine the cost imposed on the local grid. I am also interested in examining the economic synergy between solar photovoltaics and efficiency home improvements. I use data analytics and an evidence-based approach to draw policy relevant conclusions. In the field of industrial organization, I use model-based simulations to examine the potential impact of environmental policy on firm decision makings.

Name: Stephen S. Gao

Department: Geosciences and Geological and Petroleum Engineering

Key Words: Signal detection, data mining, time series analysis, seismology, earthquakes, computer-aided tomography, machine learning

A brief description of research: Pure and applied geophysics with emphasis on earthquake seismology. Seismological investigations of the Earth's crust and mantle using seismic tomography, anisotropy, and receiver function analyses. Distribution of earthquakes in space and time, crustal deformation, and earthquake hazard prediction and mitigation. Frequently utilize computing intensive analysis of time series, data mining, and machine learning techniques. Areas of study include Brazil, China, Egypt, Ethiopia, Japan, Kenya, Mongolia, Russia, South Africa, the United States, and Zimbabwe etc.

Name: Xiaoming He

Department: Mathematics and Statistics

Key words: Data-enabled modeling, data assimilation, numerical methods for partial differential equations, computational fluid dynamics

A brief overview of research: With the support of lab experiment datum, I am investigating new coupled multi-physics multi-scale models and the corresponding numerical methods for

accurately describing and simulating various complex flow problems, including the coupling of dual porosity flow and free flow, two phase ferro-fluid flows, and two-phase porous media flow coupled with two-phase free flow. D

Name: Michael G. Hilgers

Department: Business and Information Technology

Key words: Information visualization, business analytics, human-computer interaction/user experience, Neuro-IS (eye-tracking & EEG), perception model formulation and analysis, quantitative finance and data science applications, interest rate models, and hybrid options

A brief overview of research: My main research interest focuses on information visualization and its various aspects from data representation to human perception. Combinatorial constructs such as hypergraphs in conjunction with ranking and ordering are explored to tie data to the reading of a visualization. More physical models of eye motion are obtained using stochastic geometry. A different vein research proceeds from the Business Analysis and Data Science classes that I created. I am interested in pedagogical research into teaching business majors concepts in quantitative finance and related data science problems. I also pursue personal interests in interest rate models and as well as hybrid option theory.

Name: Wenqing Hu

Department: Mathematics and Statistics

Key words: probability, stochastic analysis, data science

A brief overview of research: I have been working in probability theory. By making effective use of stochastic analysis, I analyzed problems in stochastic processes, differential equations, dynamical systems, and mathematical physics. These problems include small random perturbations of dynamical systems, large deviations, metastability, stochastic averaging principle, reaction-diffusion equations and wave front propagation in random media, stochastic fluid mechanics, turbulence models, small mass limit of the Langevin equation (Smoluchowski-Kramers approximation), homogenization and multiscale problems, and system of fast-slow stochastic reaction diffusion equations. I also have interests in statistical methodology. I have been working on problems in data sciences, statistical machine learning and optimization. In particular, I got involved in covariance matrix estimation under high-Dimensional-Low-Sample-Size (HDLSS) setting, with applications to regularized linear discriminant analysis in Electronic Health Records (EHR) data, variational inference of human mobility patterns via Hawkes processes, convergence analysis of stochastic approximation algorithms (e.g. stochastic gradient descent) that are used in solving stochastic optimization problems, real-world applications of Markov Decision Processes (MDP) and Reinforcement Learning.

Name: Tony Luo

Department: Computer Science

Keywords: Internet of Things, machine learning, cyber security

A brief overview of research: I conduct research on Internet of Things security, adversarial machine learning, and AI-empowered IoT applications. My objectives are to safeguard IoT systems against vulnerabilities, make machine-learning algorithms robust to adversarial and unreliable users, and develop innovative AIoT applications that are of profound economic and societal impact.

Name: Kelly Liu

Department: Geosciences and Geological and Petroleum Engineering

Key words: Digital signal processing and analysis, geophysical modeling and inversion, subsurface imaging.

A brief overview of research: My research includes two major areas and both involve data mining and data analytics. The first area is aimed at understanding the internal structure of the Earth, how it evolves, and what causes geohazards such as earthquakes and volcanoes. The second research area is to utilize seismic data to image subsurface structure to understand depositional environment and hydrocarbon potential. Seismic attributes associated with time, amplitude, frequency, and attenuation are extracted and integrated with petrophysical information for reservoir characterization and property mapping.

Name: Sanjay Madria

Department: Computer Science

Keywords: Big data, cloud computing, data analytics, security, machine learning

A brief overview of research: My interest is in the design and building big data and cloud system infrastructure to provide efficient data management, analytics, security and learning. Several papers and systems have been designed for data analytics and security with applications in disaster management, combat cloud, IoTs, Covid-19, and others. I also have current collaborations with domain experts in the area of advanced manufacturing, steel defects detection, and chemistry in the data analytics area.

Name: Jeremy Maurer

Department: Geological Sciences and Geological and Petroleum Engineering

Key words: Remote sensing, uncertainty quantification, big data, machine learning, optimization

A brief overview of research: In our lab we focus on applying a broad range of statistical and computational techniques to solve problems in earthquake science and remote sensing. Several projects in southern California, Mexico, Haiti, and Guatemala involve using geodetic remote sensing to measure crustal strain accumulation on large faults and the resulting earthquake potential. We develop new algorithms for quantifying uncertainty in the results, since all geophysical methods are limited by the fact that we only have surface observations of what are deep sub-surface phenomena. We also study human-induced earthquakes using numerical models and statistical analysis to understand what controls the largest events. We are currently working to develop methods to correct large InSAR datasets for atmospheric noise by developing efficient computational algorithms and using machine learning to learn weather patterns.

Name: Melanie Mormile

Department: Biological Sciences

Key words: Environmental microbiology, extremophiles, halophiles, biotechnology

A brief overview of research: When most people think of microorganisms, they think of disease-causing organisms. I'm interested in using beneficial microorganisms to help clean-up contaminated environments and for industrial purposes. To gain a better understanding of such organisms, I have studied bacteria from extreme environments. My interest in data science is the handling of large amounts of genomic data to find and examine specific genes.

Name: Jason Murphy

Department: Mathematics & Statistics

Key Words: harmonic analysis, analysis of PDEs, nonlinear wave equations, inverse problems, medical imaging, nonlinear optics

A brief overview of research: To date, most of my research has focused on problems related to the long-time behavior of solutions to nonlinear dispersive partial differential equations, which refers to a class of equations arising in the modeling of various nonlinear phenomena arising in physics, especially nonlinear optics. Studying these problems requires many tools from mathematical analysis, especially harmonic analysis, functional analysis, spectral theory, etc. Recently, I have been working to learn more about problems related to the mathematics of

medical imaging, and more generally the mathematics of inverse problems, which involves many of the same mathematical tools. I am especially interested in recent developments that incorporate machine learning algorithms into the study of inverse problems related to imaging.

Name: Venkata Sriram Siddhardh Nadendla

Department: Computer Science

Key words: Cyber-Physical-Human Systems; Algorithmic Game Theory; Statistical Inference and Machine Learning; Security, Trust and Influence

A brief overview of research: My research interests and expertise lies in developing secure and trustworthy cyber-physical-human systems in the presence of competitive/strategic agents and heterogeneous resources in potentially insecure environments, especially in the context of transportation, wireless networking, mining and defense applications. Furthermore, I have also been recently working on developing strategic frameworks that promote human-system interaction and teaming in order to solve complex problems effectively, which cannot be solved by either automated systems, or humans by themselves. In solving such problems, I rely on concepts/tools in algorithmic game theory, artificial intelligence, machine learning and human decision modeling.

Name: Fiona Nah

Department: Business and Information Technology

Key words: Human-computer interaction/user experience, Neuro-IS (eye-tracking & EEG), usable privacy and security, gamification, virtual communities and collaboration, digital commerce, meta-analysis

A brief overview of research: My main research interests revolve around human-computer interaction (HCI) and user experience. I'm interested in studying the bright/positive and dark/negative sides of technology including designing interfaces for flow (or cognitive absorption) experience and studying the nature of addictive effects of technology. We are currently analyzing the EEG activity for the state of flow of users. I'm also interested to understand/develop/derive computational models for aesthetic interfaces, including the use of fractals and the golden ratio. My Master's students and I have also used and analyzed eye-tracking data for HCI research. Dr. Matt Thimgan and I are studying eye-tracking patterns and eye metrics for sleepiness. Dr. Ting Shen and I are working on a meta-analysis on gamification in the education context. I have previously received an NSF EAGER funding to understand usable security by studying how interface design and message framing can influence user behavior in the context of information security.

Name: Tayo Obafemi-Ajayi

Department: Electrical and Computer Engineering

Key words: machine learning, data analytics, clustering, biomedical applications, validation metrics, intelligent systems

A brief overview of research: My interdisciplinary research centers on developing and application of machine learning algorithms to design intelligent systems that make meaningful predictions for decision support for a variety of bioscience applications. Multiple biomedical disorders (such as Autism Spectrum Disorder, Traumatic Brain Injury (TBI), Inflammatory Bowel Disease, cognitive aging) are characterized by a high degree of heterogeneity. In such applications, unsupervised learning aids disease subtyping i.e. the task of identifying homogeneous patient subgroups that can guide prognosis, treatment decisions and possibly predict outcomes or recurrence risks. My research contributions have yielded multiple computational models to sort out the heterogeneity which includes analysis of relevance and clinical interpretability of different clustering solutions. I have collaborated with Drs. Olbricht and Wunsch on a variety of these projects. Currently for our TBI work, we have access to a wide range of data including biomarker, phenotype, brain imaging scans, collected at different time points before and after concussion from athletes and soldiers. We are developing algorithms to analyze the data to identify different severity levels of the injury, and predict recovery trajectories.

Name: Gayla Olbricht

Department: Mathematics and Statistics

Key words: Statistics/data analytics (specializing in biological/biomedical applications), machine learning, bioinformatics/statistical genomics

A brief overview of research: My general research interest lies in the statistical modeling of biological data. One of my primary application areas is in the field of bioinformatics, where data science and statistical methods are needed to analyze large, complex high-throughput datasets, enabling a better understanding of the genetic underpinnings of different traits and diseases. I have worked with various types of “omic” data including gene expression, DNA methylation, and single nucleotide polymorphism and am also interested in methods that integrate multiple “omic” data types. In addition to the bioinformatics work, I am actively working with Drs. Thimgan and Samaranayake on developing models for predicting the lifespan of fruit flies (*Drosophila*) based on sleep characteristics. I am also collaborating with Drs. Obafemi-Ajayi and Wunsch on combining statistical analysis with clustering to aid in subtyping of diseases with high heterogeneity, such as autism spectrum disorder and traumatic brain injury. Through these experiences, I have worked with a variety of statistical and machine learning methods to solve complex problems, including specific applications in Markov modeling, regression analysis, functional data analysis, survival analysis, and data imputation.

Name: Shun Saito

Department: Physics

Key words: Cosmology, Dark Energy, Galaxy evolution, Data compression, Statistical Inference

A brief overview of research: My research area lies in theoretical aspects of observational cosmology, including simulation work and data analysis. The primary goal of my research is to unveil the fundamental physics from a trove of cosmological observations. Right now I am focusing on revealing the nature of dark energy from a gigantic galaxy map. Using the 3D positions of millions of galaxies and its clustering pattern as a function of spatial scale, we try to infer model parameters (such as the dark energy equation of state) from such data as accurately and precisely as possible. A key issue is nonlinearity (or non-Gaussianity) in the data, and hence some of the machine learning techniques have been recently applied to the data analysis. I am interested in developing a new way to analyze our data, given the inputs from this group, and in applying such a method to the surveys I am working on (see e.g., our PR).

Name: V.A. (Sam) Samaranayake

Department: Mathematics and Statistics

Key words: Time Series, Prediction Models, Reliability, Statistica/data analytic applications in biology, social sciences, and engineering.

A brief overview of research: I am a statistician by training with many years of collaborative research experience with biologists, engineers, and social scientists. A major area of research interest of mine is time series analysis and prediction. I have also done research in the area of reliability analysis. In time series research, I have taken both traditional as well as novel data driven approaches to modeling. One project I am currently involved with is economic volatility modeling using new approaches. Work with Dr. Matthew Thimgan on determining the association of sleep architecture with lifespan in *Drosophila* is another project where I am using time series analysis approaches. In this latter project, hybrid models using both time series and machine learning algorithms are used. A second collaborative project with Dr. Thimgan involves identifying sleep periods of individuals based on wrist worn movement censors. In a collaborative research project with Dr. Jagannathan and his former student Krishnan Ranghava in ECE several new deep learning algorithms were developed. Other research tools I have used are functional principal component analysis, elastics nets, and the LASSO, all of which are tools for dimension reduction.

Name: Jag Sarangapani

Department: Electrical and Computer Engineering

Key words: data analytics, diagnostics/prognostics of machinery, machine learning, neural networks

A brief overview of research: My interests are in the development of novel machine learning methods to data analytics in particular diagnostics and prognostics of machinery. I was a Site Director of the NSF Industry/University Cooperative Research Center where we continue to develop novel machine learning methods for data analytics. In collaboration with Dr. Samaranayake, and my former student Krishnan Ranghava, we developed new deep learning algorithms that are mathematically rigorous and novel. We apply these machine learning based data analysis techniques to a variety of projects with machines and hospital admission prediction and so on.

Name: Daniel B. Shank

Department: Psychology

Key words: Social Psychology, Technology, Human-Computer Interaction, Artificial Agents, Morality, Social Interactions

A brief overview of research: Dr. Daniel B. Shank is an assistant professor in the Department of Psychological Science at Missouri Science & Technology specializing in the areas of social psychology and technology. He obtained a BA in Computer Science from Harding University, and from the University of Georgia he received an MS in Artificial Intelligence and an MA and PhD in Sociology. Dr. Shank served in two postdoctoral research fellowships at the University of Alabama Birmingham and at the University of Melbourne (Australia). His research interests include perceptions of and social interactions with nonhumans including artificial intelligence agents and groups of people. He studies morality, mind attributions, affective impressions, emotions, social interactions, and behavioral reactions and how these processes differ between AIs and humans and on human-AI teams. He has published over 20 articles in psychology, sociology, and behavior science technology journals and currently has grants from the Army Research Office and the Leonard Wood Institute to study affective and moral perceptions and interactions with AIs.

Name: Ting Shen

Department: Psychological Science

Keywords: Quantitative research methods, large-scale assessment data, psychometrics, empirical research in psychology and education

A brief overview of research: I am interested in applying various quantitative research methods (e.g., meta-analysis, instrumental variable, regression discontinuity design, machine learning) on large-scale data in education and psychology to inform policies and practices concerning students' overall development and wellbeing in both school and home environments. Dr. Burns and I are examining class size effects in higher education using the CET survey data at S&T. Dr. Nah and I are investigating the gamification effects on students' learning in K-12 settings.

Name: Keng Siau

Department: Business and Information Technology

Key words: Data Science, Sentiment Analytics, Emotion Analytics

A brief overview of research: In the data science area, I research in the areas of numerical analysis (data mining and predictive analysis), textual analysis (sentiment analysis), and facial expression recognition (emotion analysis). For facial expression recognition research, we use machine learning techniques as well.

Name: John Singler

Department: Mathematics and Statistics

Key words: Data driven model order reduction, proper orthogonal decomposition, singular value decomposition, data compression, partial differential equations, control theory

A brief overview of research: My research involves model order reduction and control of various mathematical models arising in applications. Model order reduction is a process of taking a complex mathematical model and creating a model of drastically lower dimension, which is useful for simulation, optimization, control design, and other applications. I have focused on model order reduction methods that utilize simulation data, and I have also performed research on data compression algorithms. My research has involved the development of new algorithms and also the mathematical analysis of the accuracy of existing algorithms.

Name: Joe Stanley

Department: Electrical and Computer Engineering

Key words: Image and signal processing; computational intelligence; data fusion; automation; bioinformatics

A brief overview of research: My multidisciplinary research collaborations include dermatology skin lesion analysis, histology image analysis for cervical cancer detection, content-based image retrieval, atomic force microscopy, x-ray image analysis for osteoarthritis assessment, intrusion detection and system survivability, and bioinformatics. In all of these research collaborations, my research focus is the investigation of image and signal processing, pattern recognition, computational intelligence, and data fusion algorithms and techniques for computer-assisted tools to aid domain experts in the decision making process.

Name: Matt Thimgan

Department: Biological Sciences

Key words: Sleep, Cognitive performance, Biomarkers, Health metrics

A brief overview of research: Our lab is interested in the molecular pathways that underlie the restorative function of sleep. We use mathematical models, designed by Dr. Olbricht and Dr. Samaranayake, to determine what aspects of sleep architecture are associated with increased lifespan in *Drosophila*. These models separate individual flies based on biological aging rather than chronological aging. Because this separation is based on sleep characteristics, we can evaluate molecular differences associated with inadequate sleep. These biochemical pathways represent potential drug targets to mitigate the deleterious effects of inadequate sleep. In humans, we are designing algorithms that can help identify physiological biomarkers of sleepiness, in collaboration with Dr. Samaranayake. We can associate metrics from tasks like balance and word pronunciation to determine if there are biomarkers of sleepiness. These metrics can be associated with cognitive performance. We typically collect repeated measures data in individuals to associate with a relevant outcome for study.

Name: Ardhendu Tripathy

Department: Computer Science

Key words: multi-armed bandits, sequential analysis, statistical learning theory, algorithms

A brief description of research: I am interested in identifying the fundamental hardness of, as well as designing new algorithms for, problems in machine learning. My solution methods typically combine perspectives from optimization, signal processing, and information theory. Some of my recent work has focused on learning from human feedback. Using sequential analysis techniques from multi-armed bandit literature, I have designed algorithms that are efficient in the amount of human feedback needed. Other work I have done is in obtaining appropriate representations of data that satisfy certain desired characteristics, such as invariance to a nuisance/sensitive attribute. Going forward, I would like to work on problems in reinforcement learning, multi-agent cooperative learning, and explainable AI.

Name: Xuerong Meggie Wen

Department: Mathematics and Statistics

Key words: Dimension reduction, Variable selection and model selection, Statistica/data analytic applications in biology, social sciences, and engineering.

A brief overview of research: My research focuses on developing new ways to extract relevant information from high dimensional data sets, while grasping the important features or patterns in the data. In some fields, such as economics, biology and finance, researchers also need to analyze high dimensional data, where the number of predictors p is frequently huge compared with the sample size n. It is often reasonable to assume only a small number of predictors actually contribute to the response. Estimation accuracy and model interpretability can be greatly improved in the subsequent analysis by effectively identifying the few important predictors first. Hence, dimension reduction or feature selection is often conducted as the first step of data analysis. Many methods have been developed in recent years to extract the significant variables effectively under the so-called $n<<p$ context. However, most of the popular variable selection methods are model based, where a linear model or generalized linear model is assumed. Such methods might generate biased results if the underlying modeling assumption is violated, which is typically the case for complex or unknown models. My recent research interests center around the development and application of model-free dimension reduction and variable selections (screening) methods in answer to the demand.

Name: Donald Wunsch

Department: Electrical and Computer Engineering

Key words: Neural networks, unsupervised learning, reinforcement learning, lifelong learning, fuzzy systems, evolutionary computation, biomedical, infrared

A brief overview of research:
I enjoy all aspects of machine learning research, particularly involving neural networks. My main contributions have been in unsupervised learning and reinforcement learning. These are prominent tools in my current projects, which include: continuous learning methods known as lifelong learning, the related topic of clustering of streaming data, validation and explanation of learned models, energy research, explainability and ethics of machine learning, biomedical applications such as prognosis of concussion recovery, infrared image analysis, and autonomous robot applications.

Key Words: Name: Yanzhi Zhang

Department: Mathematics and Statistics

Numerical differential equations, mathematical modeling, numerical algorithm and analysis, machine learning, data-driven modeling, inverse problems

A brief description of research: I am in the area of computational and applied mathematics. My general interest includes developing mathematical models, designing numerical algorithms, and studying real-application problems. Some of my ongoing research include: (1) data-driven modeling and inversion of seismic data; (2) machine learning algorithms for pattern recognition; (3) simulations on Bose-Einstein superfluids; (4) simulations and applications of nonlocal/fractional models.