NJIT | YING WU COLLEGE OF COMPUTING

RESEARCH REPORT 2025

Table of contents

YING WU COLLEGE OF COMPUTING

The Ying Wu College of Computing (YWCC) stands at the forefront of computing research, driving innovation and shaping the future of technology. Our faculty and students are pioneering breakthroughs in artificial intelligence, cybersecurity, data science, human-computer interaction and beyond—advancing both theoretical foundations and real-world applications.

As an R1 research institution, we are committed to high-impact scholarship. Our faculty are leading innovative, interdisciplinary research that tackles some of the most critical challenges facing society today: enabling the creation of smart, sustainable cities and intelligent infrastructure through mobile sensing and federated machine learning; improving physician training and health care outcomes through innovations in virtual, augmented and extended reality; and advancing understanding of solar weather and the prediction of solar flares through innovations in AI and cyberinfrastructure.

At the Ying Wu College of Computing, excellence in research isn't something reserved for the few—it's a core part of our culture that benefits every student. And as the largest producer of computing professionals in the greater New York metro area, our college is a powerhouse of discovery, creativity and impact.

This report highlights the depth and breadth of our research portfolio. I invite you to explore the high-impact projects within these pages and encourage you to connect with our researchers. We welcome new ideas, interdisciplinary collaborations and industry partnerships that push the boundaries of computing and shape a more intelligent, secure and connected world.

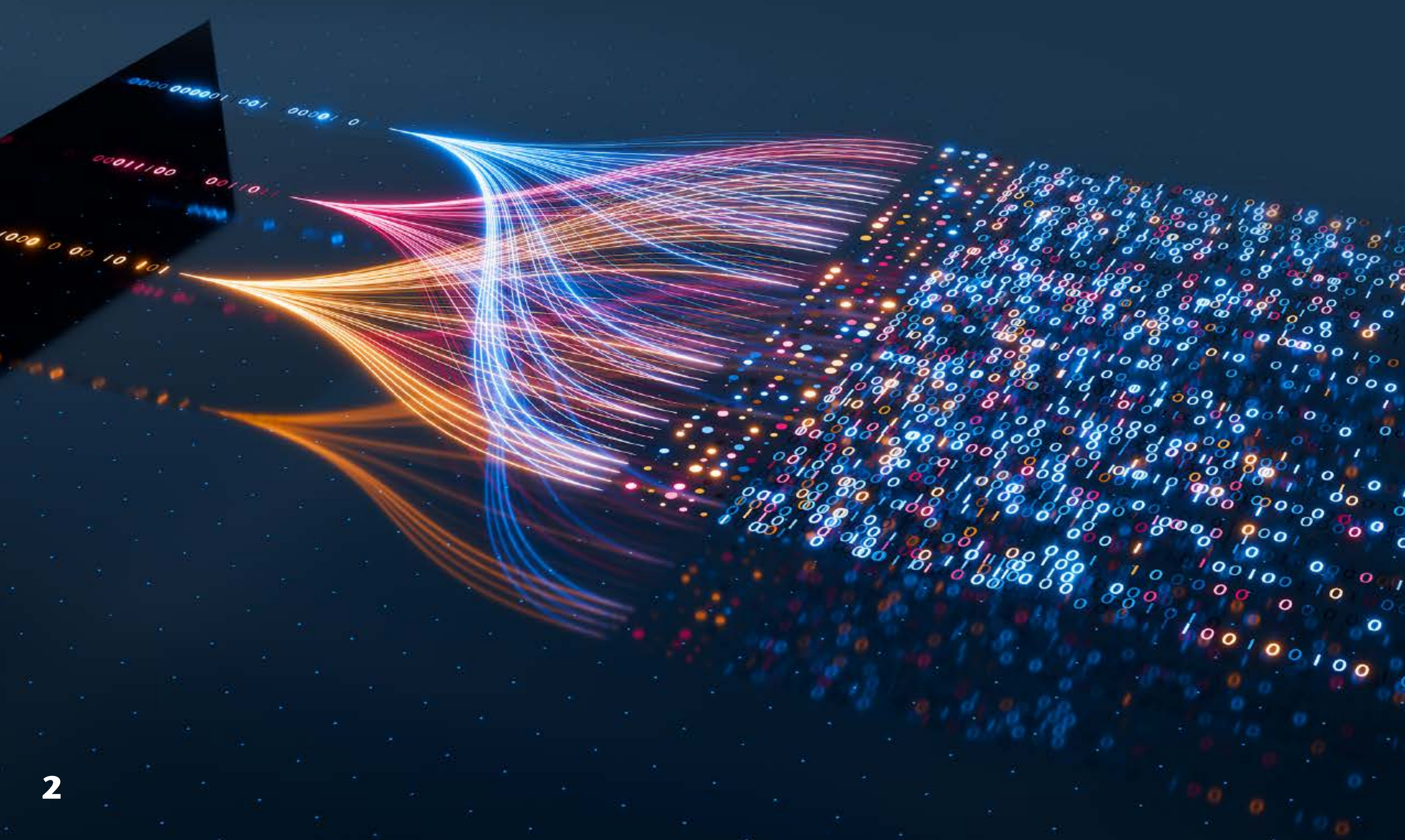
Together, let's drive the future of computing.

Sincerely,

Jamie Payton
Dean, Ying Wu College of Computing
New Jersey Institute of Technology

Advanced Algorithms

We build strong foundations in algorithms with proven performance. Our work tackles challenging problems in global optimization for non-convex tasks, spectral graph theory and exact solutions for complex computational issues. By blending theory with real-world applications, our research contributes to fields like electronic design, fair decision-making, energy-efficient resource allocation and risk assessment. We also address societal challenges, such as ensuring fairness in rideshare services. Across all projects, we rely on rigorous mathematical analysis to ensure our solutions are both theoretically sound and practically useful.



James Calvin
Professor
james.m.calvin@njit.edu

Research Area: Global optimization, probabilistic analysis of algorithms

Global Optimization

Optimization is a basic component of many algorithms for problems such as image registration, data clustering and other machine learning models. Such problems usually give rise to objective functions that have many local minima and so can not be solved with standard convex optimization methods. We develop global optimization algorithms that approximate the minimum value of such functions. Worst-case error bounds are not useful in this setting and so we establish average-case lower and upper bounds under certain probability models.

Ioannis Koutis
Associate Professor
ioannis.koutis@njit.edu

Research Area: Fast Linear system solvers, spectral graph algorithms, machine learning, data mining of algorithms

Spectral Graph Algorithms

Graphs are objects of central interest in data science. A graph can be mapped to linear operators whose spectral properties encode connectivity information, enabling the design of numerical algorithms for various problems on graphs. The practical applicability of such algorithms hinges on the existence of fast solvers for fundamental computational problems, such as systems of linear equations and other generalized regression problems. We have designed such solvers and leveraged them in the design of new graph algorithms based on spectral graph theory for graphs with a prescribed cut structure. As a concrete application, we developed new algorithms for hypergraph partitioning, a key problem in Electronic Design Automation. Our algorithms broke previous records on multiple benchmarks and received the best paper award at the 41st ACM/IEEE ICCAD, in 2022.

Exact Algorithms

Multiple well-studied computational problems are conjectured to require exponential time for their solution. Current research aims to develop a detailed understanding of the computational complexity of these problems beyond the classical NP- completeness theory. The design of faster exact algorithms for such problems and their parameterized versions is of key importance in the area. In this context, we pioneered the general method of algebraic fingerprints that reduces various combinatorial problems to monomial detection problems that are, in turn, solved via algebraic algorithms. This research led to breakthrough results for classical algorithmic problems, such as the Hamiltonian cycle problem and single exponential time ,algorithms for problems parameterized by treewidth. Among these breakthrough algorithms, our work on Directed Hamiltonicity and Outbranching problems received the best paper award at the 44th ICALP in 2017.

Marvin Nakayama
Professor
marvin@njit.edu

Research Area: Fast Linear system solvers, spectral graph algorithms, machine learning, data mining

Efficient Computational Methods for Risk Assessment

Many disparate fields suffer from uncertainties with detrimental consequences, such as large losses in financial portfolios or failures of critical infrastructure due to natural disasters. Modern society crucially depends on gaining a better understanding of the likelihood and impacts of such calamitous rare events. Our work devises novel approaches for substantially reducing statistical errors in Monte Carlo simulation, a computational technique that can be employed to study risks in decision-making and analytics. One project focuses on designing improved methods for probabilistic safety assessments of nuclear power plants. The work also applies to evaluating the reliability of complex systems such as aircraft navigation computers, package-tracking systems for overnight delivery companies and the dependability of supply chains.

Baruch Schieber
Professor
sbar@njit.edu

Research Area: Advanced algorithms, combinatorial optimization, graph theory, scheduling

Fair preference aggregation

Preference queries leverage different preference aggregation methods to aggregate individual preferences and come up with a single output. These queries are prevalent in high fidelity applications, including search, ranking and recommendation, hiring and admission and electoral voting systems. Frequently, the preference aggregation needs to obey fairness constraints that ensure proportionate representation of multiple groups based on protected attributes, such as gender or race. We design algorithms for preference aggregation that satisfy fairness constraints. We focus on two design approaches. In the first approach, we define a metric on the possible outcomes of the preference aggregation and find the outcome that satisfies the fairness constraint and is also closest (with respect to the defined metric) to the best unconstrained outcome. In the second approach, we define a metric on the possible inputs and find the smallest amount of change in the input (with respect to the defined metric) that will guarantee that the best unconstrained outcome satisfies the fairness constraints. It turns out that most of the resulting computational problems are hard and thus we resort to either heuristics or approximations with proven bounded performance.

Energy Harvesting in Resource Allocation

Conserving reusable resources and harvesting renewable resources have become critical goals in a world facing dramatic environmental and climate changes. Energy harvesting (also known as energy scavenging) is the conversion of ambient energy present in the environment into electrical energy. Energy harvesting can be used to provide either an alternative or an augmentation to batteries. Energy harvesting enables the ability to obtain renewable resources while operating a system or servicing customers, albeit at the cost of suspending regular activities for certain time intervals. We study the theory of sustainable resource allocation with energy harvesting. We define new theoretical models that correspond to advanced new technologies and capture the combination of sustainability and traditional objectives. We also study and analyze optimization problems arising in these models. Our work on interweaving real-time jobs with energy harvesting to maximize throughput received the best paper award at the 17th WALCOM in 2023.

Pan Xu
Assistant Professor
pan.xu@njit.edu

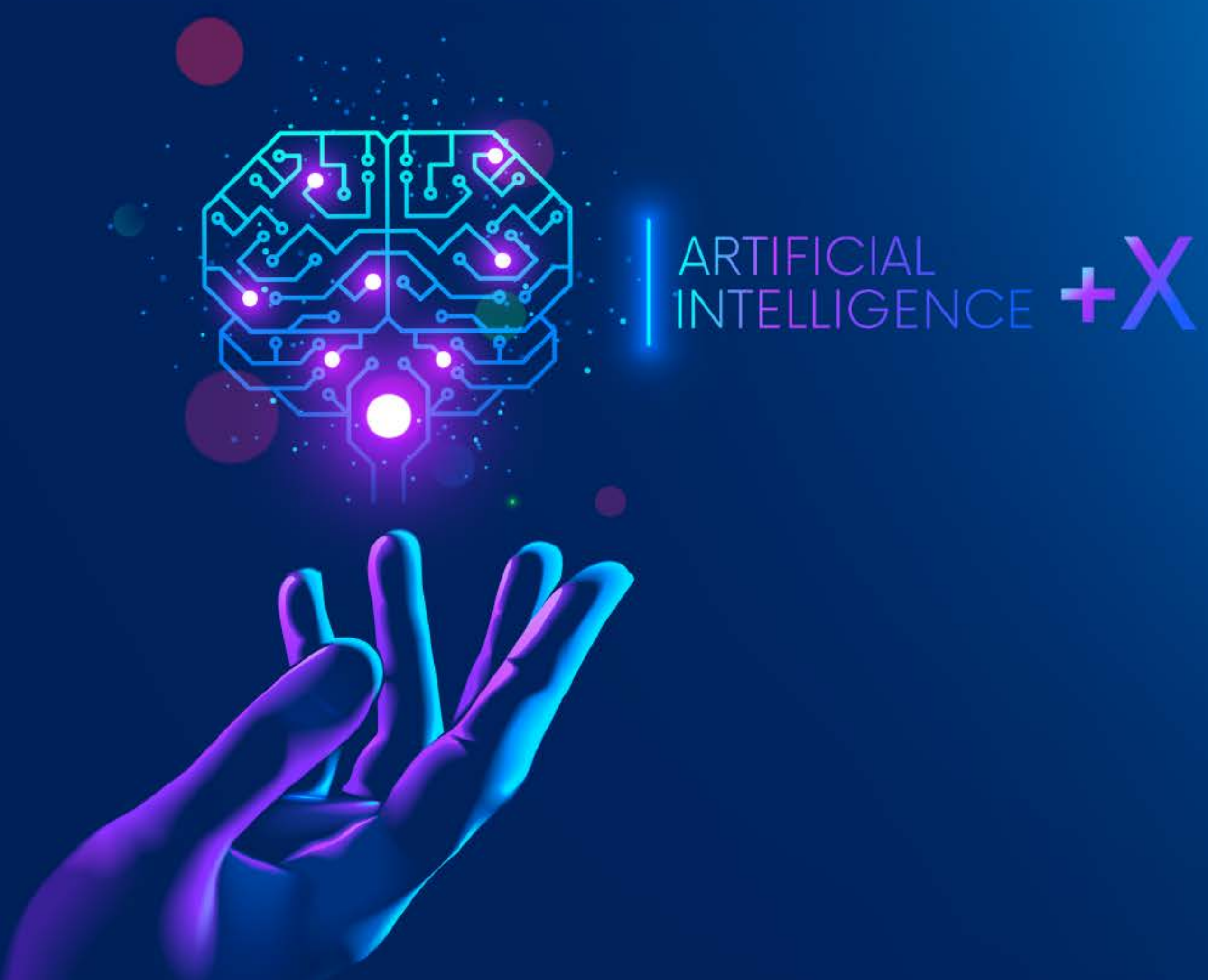
Research Area: Artificial intelligence, approximation and randomized algorithm, ridershare and crowdsourcing markets

Algorithm Design to Combat Discriminations in Rideshare Services

Rideshare platforms, such as Uber and Lyft, have gained increasing popularity in recent years. One of the central tasks facing Uber and Lyft is the matching policy pairing drivers and riders. Recently, it has been reported that drivers cancel riders based on their demographic attributes such as gender, race and disability, either intentionally or unintentionally. In our project, we try to leverage the power of algorithm design to curb discriminative cancellations from drivers to riders and improve the social welfare overall.

AI + X

AI + X signifies the interdisciplinary application of artificial intelligence (AI) to a diverse array of fields. We are exploring innovative applications of AI, such as machine listening for audio and music analysis (AI + Audio, AI + Music), federated learning for privacy-preserving AI (AI + Cybersecurity), financial market prediction and portfolio management (AI + Finance), solar astronomical data generation (AI + Astronomy) and false claim detection (AI + Journalism), as well as investigating the societal implications of AI (AI + Ethics). By combining AI's power with domain-specific knowledge, we are advancing both AI and addressing critical challenges across various disciplines. Our cutting-edge work is consistently published at both leading AI and domain-specific venues and its impact is amplified via partnerships with industry for technology transfer, such as NEC, Accenture, Adobe and Qualcomm.



Mark Cartwright
Assistant Professor
mark.cartwright@njit.edu

Research Area: Machine listening, interactive machine learning, human-computer interaction, audio processing, music information retrieval, sound accessibility

Machine Listening with Limited Labeled Data

Progress in machine listening has historically been hindered by limited amounts of labeled data. This is due to a variety of reasons: we don't share and tag audio as we do images; licensing terms prevent the distribution of annotated musical audio; and, unlike with images, the average person cannot effectively label audio with just a glance and a click — it takes time to listen. To address this, we have developed and studied methods for self-supervised audio representation learning, few-shot sound event detection, crowdsourced audio annotation and active learning.

Interactive Open-World Machine Listening

Smart acoustic sensing powered by machine listening has the potential to provide valuable information for numerous tasks, including understanding and improving the health of our cities (e.g., monitoring and mitigating noise pollution) and natural environments (e.g., monitoring and conserving biodiversity). However, contemporary machine listening models can only detect and describe a small fraction of the sounds we care about in the world. To the "ears" of these models, unknown sound classes do not exist or are confused with known classes. This results in a limited view of the acoustic world by sound event detection models that may not align to the goals of end users, hindering the machine listening's transformative potential. We aim to develop methods and tools to detect, discover, describe and define "unknown" sound events in support of large-scale longitudinal audio analysis and model development.

Guiling "Grace" Wang
Distinguished Professor
gwang@njit.edu

Research Area: Deep learning, AI in Finance, AI in Transportation, LLM.

Financial Market Prediction by Leveraging Generative Adversarial Networks and Graph Neural Networks

Financial market prediction faces several key challenges: (1) the stochastic nature of markets requires models to quantify uncertainty, yet most point prediction methods provide limited information; (2) traditional models often overlook market risk; and (3) current models focus on static asset relationships, ignoring the evolving dynamics of financial networks. To address these issues, we first propose RAGIC, a Risk-Aware Generative model for Interval Construction, which focuses on interval prediction by forecasting price ranges that better capture market uncertainty. It also incorporates the volatility index to detect market fluctuations and adaptively adjust interval width in response to risks. In addition, DySTAGE, a novel dynamic graph representation learning framework for asset pricing is introduced. It contains a universal formulation accommodating changing asset composition and connections and a graph learning model capturing both topological and temporal patterns enriched with financial insights, making it highly adaptable to dynamic market conditions. These models offer a more robust and flexible approach to financial forecasting.

Portfolio Management by Utilizing Deep Reinforcement Learning and Large Language Models

Existing approaches in portfolio management mainly focus on cash-only trading, overlooking potential benefits and risks associated with margin trading, particularly in short sale scenarios. To address it, Margin Trader, an innovative RL framework, integrates margin accounts and constraints into a realistic trading environment, supporting both long and short positions. This framework not only maximizes profits but also manages risks effectively in both bullish and bearish markets, offering traders the flexibility to customize settings in line with market conditions, risk tolerance and personal strategies. Additionally, to dynamically adjust portfolios between long and short positions in response to evolving market conditions, LLMs are incorporated to learn market trends from diverse external data sources, from financial time series to news, enabling a more explainable and transparent decision-making process with a clear reasoning path. By combining RL with the reasoning capabilities of LLMs, this approach significantly enhances market forecasting, strategy optimization and overall performance in portfolio management.

Jason Wang
Professor
wangj@njit.edu

Research Area: Data mining, machine learning, deep learning, explainable AI, generative AI, trustworthy AI, data science

Understanding Solar Astronomy with Generative AI

Many magnetic field parameters related to solar eruptions, including flares and coronal mass ejections, are derived from vector magnetograms in the Sun's atmosphere. However, high-resolution high-cadence time series vector magnetograms are lacking in previous solar cycles (except solar cycle 24). In this project, we employ generative AI techniques to create synthetic vector magnetograms in previous solar cycles, which will enable new discoveries in solar astronomy. In addition, these generative AI techniques help (1) enhance spatial and temporal resolutions of observations taken by space-borne and ground-based instruments, (2) clean the observations by removing noises from them and (3) create synthetic EUV images that can provide crucial information of potential radiation hazards such as radio blackouts. The project showcases many important applications of generative AI in astronomy, space physics and solar science.

Brook Wu
Associate Professor
wu@njit.edu

Research Area: Applications of knowledge representation in fake news detection, knowledge discovery and improving student learning outcomes

The 'Path' to Clarity: Identifying False Claims Through a Knowledge Graph Exploration

Wu's current research focuses on advancing automated fact-checking by addressing key challenges: claim ambiguity, lack of contextual information, evolving facts and claim complexity. A key theme across this work is the use of Knowledge Paths (KPs) retrieved from knowledge graphs to provide additional context and disambiguate claims. These KPs enhance the representation of both claims and evidence, improving the model's ability to detect falsehoods. Our work also introduces advanced graph-based neural networks to capture semantic relationships between claims and evidence. Additionally, it addresses the issue of evolving facts, where knowledge changes over time. Our models adapt by incorporating newly available information, improving the robustness of fact-checking systems in dynamic environments. An extension of this work also explores a question-driven approach that leverages LLMs to systematically disambiguate and decompose complex claims into simpler, verifiable subclaims, resulting in more accurate verification processes. Across these contributions, the research emphasizes efficiency by dynamically retrieving evidence only when necessary, ensuring that external knowledge is incorporated without redundant processes.

Yvette Wohn
Associate Professor
wohn@njit.edu

Research Area: AI ethics, human-computer interaction

Responsible AI

This research looks at the ethical and societal implications of using artificial intelligence in social domains that impact our everyday life, such as the detection of hate speech in social media, use of generative AI in creative domains such as art and music and the effect of AI conversational agents on mental health/well-being. We also look at how AI can help and hinder marginalized populations.

Big Data

We conduct research on a wide range of topics that span multiple layers of the technology stack within the big data ecosystem. Our work covers everything from big data transfer over high-performance networks, distributed data storage and management and parallel computing and processing, to green computing and networking, large-scale workflows, massive-scale graphs, high-performance data analytics and dimensionality reduction. Our research develops and integrates novel algorithms and techniques aimed at improving and optimizing the performance of big data systems and applications. Our work has been featured at top-tier venues, including VLDB, SIGMOD, WWW, KDD, IPDPS, SC, INFOCOM and eScience, with multiple best paper awards at various international conferences.

David Bader
Distinguished Professor
bader@njit.edu

Research Area: Data science, high performance computing, cybersecurity, computational science and engineering

Cyber-Infrastructure for Community Detection, Extraction and Search in Large Networks

Networks, or graphs, consist of nodes (vertices) connected by edges (links) and are used to represent various types of information such as social relationships or biological interactions. Community detection, also called graph clustering, is a key problem that involves grouping vertices into cohesive clusters based on their connections. This project focuses on developing new, efficient methods for community detection that can handle very large networks with millions or billions of nodes. The research will address limitations in current approaches, like poorly connected clusters and introduce the Connectivity Modifier software, which improves cluster quality. The project aims to create a modular suite of clustering tools, optimized for high-performance computing platforms. Additionally, it includes educational initiatives, providing advanced training for undergraduate and graduate students. The outcomes are expected to benefit diverse fields such as biology, engineering and social sciences by enabling analysis of massive networks.

Vincent Oria
Professor
oria@njit.edu

Research Area: Multimedia databases, spatio-temporal databases, recommender systems

Dimensionality and Scalability Issues in High-Dimensional Spaces

When researching fundamental operations in areas such as search and retrieval, data mining, machine learning, multimedia, recommendation systems and bioinformatics, the efficiency and effectiveness of implementations depends crucially on the interplay between measures of data similarity and the features by which data objects are represented. When the number of features known as data dimensionality is high, the discriminative ability of similarity measures diminishes to the point where methods that depend on them lose their effectiveness. Our research looks at the interplay between local features, the intrinsic dimensionality and their application to search, indexing and machine learning.

Multi-Instrument Database of Solar Flares

Solar flares are the most prominent manifestation of the sun's magnetic activity. They emit radiation that could potentially damage power systems, interfere with civilian and military radio frequencies and disrupt spacecraft operations. To improve analysis, in collaboration with the department of physics, we aim to integrate, clean and enrich solar data captured by various solar flare observing instruments around the world and are using them for some predictive analysis tasks.

Senjuti Basu Roy
Associate Professor
senjuti.basuroy@njit.edu

Research Area: Human-in-the-loop large-scale data analytics, optimization algorithms

Big Data Analytics Laboratory - Data Analytics with Humans-in-the-Loop

The Big Data Analytics Lab (BDaL) is an interdisciplinary research laboratory that focuses on large-scale data analytics problems rising in different application domains and disciplines. One focus of our lab is to investigate an alternative computational paradigm that involves humans-in-the-loop for big data. These problems arise at different stages in a traditional data science pipeline, such as data cleaning, query answering, ad-hoc data exploration or predictive modeling, as well as from emerging applications. We study optimization opportunities that arise because of this unique man-machine collaboration and address data management and computational challenges. Our focus application domains are crowdsourcing, social networks, health care, climate science, retail and business, naval applications and spatial data.

Chase Wu
Professor
chase.wu@njit.edu

Research Area: Big data, machine learning, green computing and networking, parallel and distributed computing

Revolutionizing Big Data Scientific Computations

Next-generation scientific applications are experiencing a rapid transition from traditional experiment-based methodologies to large-scale simulations featuring complex numerical modeling with a large number of tunable parameters. Such model-based simulations generate colossal amounts of data, which are then processed and analyzed against experimental or observation data for parameter calibration and model validation. The sheer volume and complexity of such data, the large model-parameter space and the intensive computation make it practically infeasible for domain experts to manually configure and tune hyperparameters for accurate modeling in complex and distributed computing environments. We develop visualization algorithms for 3D volume data generated by scientific computations on supercomputers and apply machine learning techniques to automate, expedite and optimize the parameter tuning process in model development.

Modeling and Optimizing Big Data Ecosystems

The execution of big data workflows is now commonly supported on reliable and scalable data storage and computing platforms such as Hadoop. There are a variety of factors affecting workflow performance across multiple layers in the technology stack of big data ecosystems. Modeling and optimizing the performance of big data workflows is challenging because the compound effects of such technology layers are complex and opaque to end users. We develop a cross-layer coupled design framework, which integrates information theory-based feature selection and stochastic approximation-based profiling to automate and optimize the configuration of big data ecosystems.

Optimizing Distributed Training and Inference of Deep Neural Networks (DNNs)

Deep Neural Networks (DNNs) have grown rapidly in size and complexity, requiring various data/model/tensor parallelization techniques to make training/inference practically feasible. For example, BLOOM 176B and Megatron-Turing 530B require terabytes of memory and zettaflops of compute. We represent parallelized DNNs as workflows and develop new approaches to workflow partitioning, mapping and scheduling alongside memory saving techniques such as activation recomputation to optimize the training and inference processes of DNNs in heterogeneous multi-node, multi-GPU/CPU systems.

Reducing Energy Consumption in Big Data Computation

The transfer of big data across high-performance networks consumes a significant amount of energy. Employing two widely adopted power models — power-down and speed scaling — we have made inroads into green computing and networking in big data environments. Our approach allows network providers to reduce operational costs and reduce carbon dioxide emissions.

Uncovering Low-Level, Hazardous Radiation

Radioactive substances and biological agents present a serious threat to public health and safety, particularly in densely populated areas. Through the collection and analysis of large amounts of sensor measurements, we develop reliable tools to detect and contain radioactive materials to protect the populace and reduce the risk of radiological dispersal devices, such as so-called dirty bombs.

Bioinformatics and Medical Informatics

Our research is focused on advancing machine learning, statistical modeling, natural language processing and data analytics to tackle key challenges in biomedical fields. We work on developing novel deep learning methods for analyzing single-cell genomic data, improving medical informatics and creating medical ontologies. Recent projects include modeling Social and Commercial Determinants of Health, which significantly influence public health and building ontologies to better understand their impacts. Our work also applies deep learning to medical AI challenges, such as tumor identification in brain MRI images and vessel identification in vascular ultrasound and enhances electronic health records (EHR) usability through machine learning. This includes improving comprehension of clinical notes with large language models and other methods to simplify EHR text for patients. In addition, we are leveraging generative models to predict Alzheimer's disease progression by addressing missing data issues and under-representation in marginalized populations. Our aim is to integrate complex, multimodal data for more accurate prediction models in aging-related diseases. Overall, our efforts focus on integrating biological information and computational techniques to improve both the analytical performance and interpretability of complex biomedical data.

James Geller
Professor
geller@njit.edu

Research Area: Medical Informatics, Medical Ontologies/Terminologies

Social and Commercial Determinants of Health

The pivotal impact of Social Determinants of Health (SDoH) on people's health and well-being has been widely recognized and researched. These include upbringing, living and working environment and access to transportation and healthcare. In simple words, it is easier to stay healthy if you have financial means. However, the effect of Commercial Determinants of Health (CDoH) is only now garnering increased attention. The study of CDoH can offer a systematic approach to identifying and categorizing the diverse commercial factors affecting health, e.g., advertising of tobacco products. These factors, including the production, distribution and marketing of goods and services, may exert a substantial influence on health outcomes. An ontology is a well-organized collection of the major terms in a subject area. We have developed an ontology for CDoH by utilizing published medical articles and ChatGPT and an ontology for SDoH and integrated them.

Yehoshua Perl
Professor
perl@njit.edu

Research Area: Medical informatics, Semantic web, ontologies interface terminologies, quality assurance of biomedical terminologies, electronic health records

Highlighting of Electronic Health Records (EHR)

We use Machine Learning to design an interface terminology to highlight the important content of EHR notes for fast skimming.

Simplification of EHR notes

It will enable patients to comprehend their clinical notes in Portals. EHR texts are difficult. We use highlighting, LLM summaries and ML techniques.

Usman Roshan
Associate Professor
usman.w.roshan@njit.edu

Research Area: Machine learning, medical informatics

Deep Learning for Medical AI Problems

We work on several medical AI problems involving images and DNA. In collaboration with neuroscientists, we are creating new models for tumor identification in brain MRI images and are studying them across different data sources. With vascular surgeons we are proposing novel models for identifying vessel and plaque in vascular ultrasound images from real patients. We are studying simple random networks for the classification of histopathology slide images and find them to be highly accurate there.

Zhi Wei
Distinguished Professor
zhiwei@njit.edu

Research Area: Deep learning, statistical modeling, bioinformatics

Deep Learning Methods to Integrate Biological Information for Analysis of Single-cell RNAseq Data

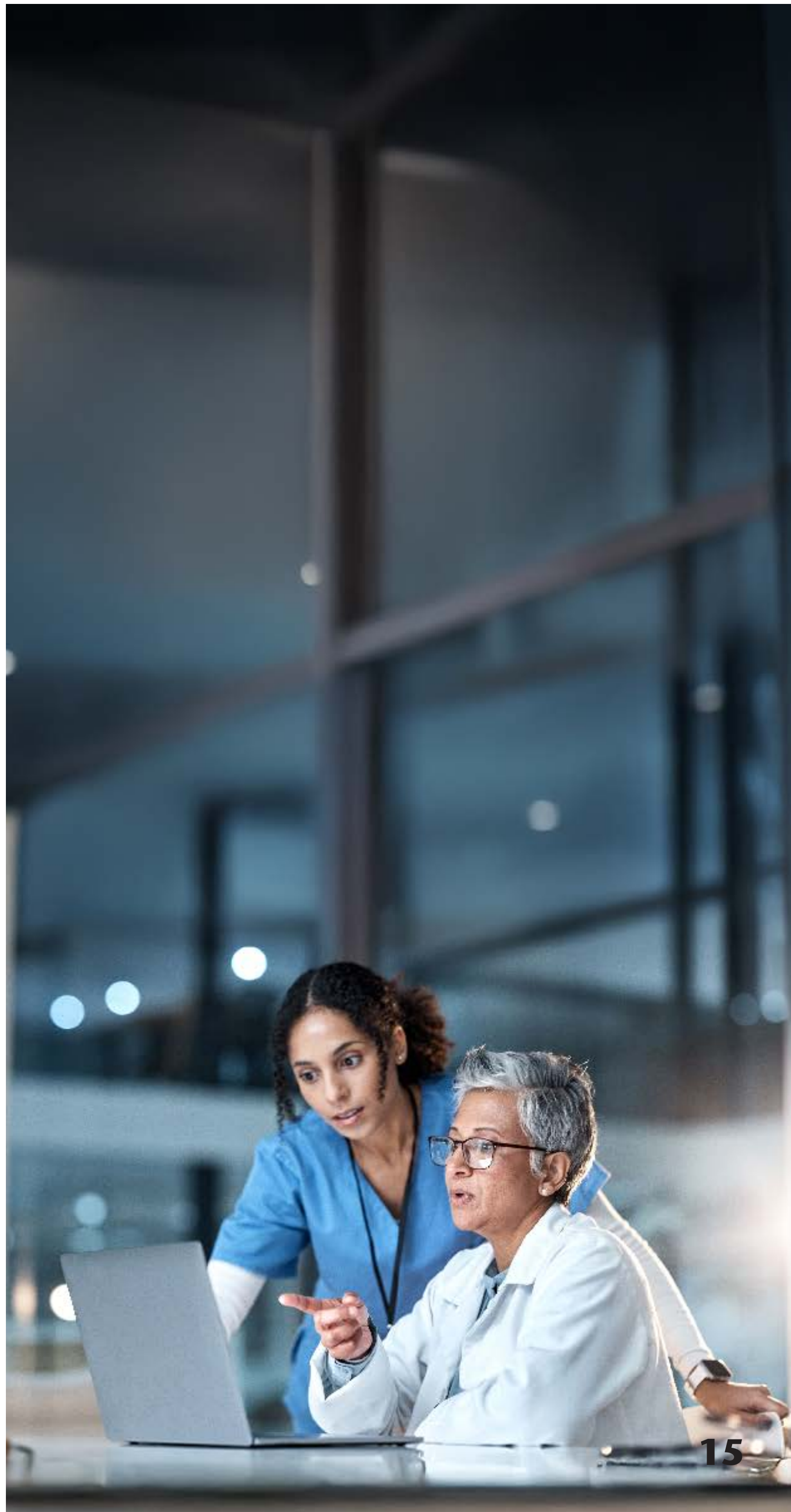
The broad long-term objective of the project concerns the development of novel machine learning methods and computational tools for modeling genomic data, motivated by important biological questions and experiments. The analysis of single-cell RNA sequencing (scRNAseq) data presents substantial computational and bioinformatics challenges. The specific aim of the project is to develop novel model-based deep learning methods, with prior biological information considered, for modeling scRNAseq data. These problems are driven by collaborations with biomedical investigators and the proposed approaches are designed to integrate biological information to improve both analytical performance and biological interpretability. Central to the project is the novel integration of biological insights with advanced deep learning methods for analyzing the noisy, sparse and over-dispersed scRNAseq data. These methods include the zero-inflated negative binomial model, autoencoder, deep embedding, hyperbolic embedding and reversed graph embedding. Additionally, the project will explore leveraging large language models (LLMs) specifically for single-cell omics data. LLMs will be applied to improve the modeling of complex gene expression patterns, capture biological heterogeneity and assist in interpreting the high-dimensional data characteristic of single-cell omics. By combining LLMs with deep learning techniques, we aim to create more robust computational models that enhance biological discovery and provide deeper insights into cellular processes.

Chenxi Yuan
Assistant Professor
cy324@njit.edu

Research Area: Deep learning, generative AI, bioinformatics

Enhancing Alzheimer's Disease Progression with Deep Learning Models

This project aspires to refine the predictive modeling of Alzheimer's disease (AD) progression by leveraging multimodal longitudinal data, encompassing biomarkers, neuropsychiatric results and neuroimaging. A significant obstacle in such studies is the lack of comprehensive data due to small sample sizes, limited data elements and the issue of missing data. This is particularly concerning for marginalized populations, either affected by AD or from ethnic minority groups. Existing imputation techniques inadequately address the intricate patterns in high-dimensional data. This project aims to counter these challenges by developing generative AI models to impute missingness and augment data for under represented groups with simulated data to enhance fairness. Long-term, this initiative seeks to develop an end-to-end prediction system that will accelerate prediction modeling for aging-related problems addressing the challenges of missing data, under representation and integration of complex, multimodal data.



Blockchain Technologies

Blockchain technology has attracted increasing attention from both academia and industry. Its popular applications include cryptocurrency, smart contracts, non-fungible tokens (NFT), Web3 and many others. It is an interdisciplinary field encompassing distributed systems, security, privacy, cryptography, database and compiler. We conduct fundamental and applied research in blockchain consensus and decentralized applications. Our projects are funded by federal agencies, such as FHWA and the results of our research are published in top venues, such as IEEE Transactions on Parallel and Distributed Systems and IEEE ICDCS.

Guiling "Grace" Wang
Distinguished Professor
gwang@njit.edu

Research Area: Deep learning, FinTech, blockchain technologies, intelligent transportation systems

Decentralized Vehicle Credential Management System Based on Consortium Blockchain

To realize the full potential of connected and automated vehicles to transform the US transportation system, it is important to ensure the vehicle communication is secure and privacy-preserving. In this situation, a digital certificate is critical to provide communication integrity, authenticity and privacy. Existing solutions based on centralized security credential management systems are inherently vulnerable to emerging cyber-security risk and single-point failure. To address the weakness of centralized systems being adopted, this project seeks to utilize the recent breakthrough of blockchain technology, in particular the consortium blockchain, to design and implement an innovative decentralized vehicle credential management system. The proposed decentralized credential management system based on consortium blockchain can not only greatly improve the robustness and security of the credential management system for vehicles and thus realizes high-security assurance for authentic V2V and V2I communication, but also incorporate state-of-the-art achievements in cryptography to ensure strong anonymity to prevent hackers from tracing the locations of vehicles and ensuring vehicle privacy.

Chase Wu
Professor
chase.wu@njit.edu

Research Area: Big data, machine learning, green computing and networking, parallel and distributed computing

A Blockchain-Based Decentralized IIoT Management System

Blockchain has been increasingly used to secure data collection, storage and transfer in various Industrial Internet of Things (IIoT) environments. We design a blockchain-based decentralized IIoT management system for censorship resistance, which employs a diffusion mechanism to deliver messages from sensors to full nodes, an augmented consensus protocol to check data losses and facilitate opportunistic outcome delivery and a public key-based aggregation method to reduce communication complexity and signature verification.

Cloud/ Edge/High-Performance Computing

Cloud computing, edge computing and high-performance computing (HPC) are driving forces behind innovations in fields such as AI, machine learning and real-time systems. Cloud platforms provide scalable, on-demand resources, while edge computing processes data closer to its source, reducing latency and offloading computational burden from the centralized cloud. HPC maximizes computational power to tackle large-scale, complex problems through dedicated hardware and special software. Typical infrastructure in these areas consists of servers connected by high-speed networks, each equipped with high-end CPUs and GPUs. In the post-Moore's Law era, the convergence of these areas, with increasingly heterogeneous, disaggregated and dynamic resources, presents both opportunities and challenges. Innovations in resource management, parallel computing and system design are now essential to fully harness this evolving landscape, creating the foundation for next-generation computing infrastructures and applications.

Xiaoning Ding
Associate Professor
xiaoning.ding@njit.edu

Research Area: Cloud/edge computing infrastructures, system designs for AI/DL, system software designs, database and data storage systems

Management of Disaggregated and Dynamic Resources in Clouds and Edges

In the post Moore's law era, computing resources are undergoing fundamental changes in many aspects (e.g., types, architectures and features). In clouds and edges, computing resources are increasingly heterogeneous (e.g., many varieties of processors and accelerators), disaggregated (e.g., local and remote memory pooled together and made available through fast network) and dynamic (e.g., resource availability changing over time). These changes enable new computing paradigms and optimization opportunities yet raise new challenges in resource management.

System Software for Scalable Computation in the Cloud

As computational resources continue to increase, we need ways to scale the performance of these computers by taking advantage of the extra resources. The objective is to guarantee that applications in the cloud can achieve higher performance when presented with more resources.

Alex Gerbessiotis
Associate Professor
alexg@njit.edu

Research Area: Architecture-independent parallel algorithm design and implementation

Multi-Core and Many-Core Algorithm Design, Analysis and Implementation

We study models of computation that abstract and capture parallelism in the presence of multiple memory hierarchies and cores. New approaches are needed to make multi-core architectures accessible to software designers in domains such as machine learning and big data. Abstracting the programming requirements of such architectures in a useful and usable manner is necessary to increase processing speed and improve memory performance.

Parallel Computing Techniques in Sequential Serial Computing

The norm in computing is to port sequential algorithms that work on one processor into multi-core or parallel algorithms intended for multiple cores and processors. Amdahl's Law highlights the limitations of using multiple cores in programs with an inherently sequential component that is not amenable to parallelization. We address this by exploring the utilization of parallel computing techniques to speed up a sequential program by exploiting the multiple memory hierarchies present in contemporary microprocessors, even if its multi-core capabilities are left unexploited.

Jing Li
Associate Professor
jingli@njit.edu

Research Area: Real-time systems, parallel computing, cyber-physical systems and reinforcement learning for system design and optimization

Parallel Real-Time Systems

Real-time systems need to provide timing guarantees for latency-critical applications in cyber- physical systems that interact with humans or the physical environment. Examples span autonomous vehicles, drones, avionic systems and robotics to structural health monitoring systems and hybrid simulation systems in earthquake engineering. However, as parallel machines become ubiquitous, we face challenges in designing real-time systems that can fully utilize the efficiencies of parallel and heterogeneous computing platforms. We are developing parallel real-time systems by exploiting the untapped efficiencies in the parallel platforms, drastically improving the system performance of a cyber-physical system.

Scheduling for Interactive Cloud Services

Delivering consistent interactive latencies, such as response delays, is the key performance metric of interactive cloud services that significantly impacts user experience. The need to guarantee low-service latency, while supporting increasing computational demands due to complex functions of the services, requires parallel scheduling infrastructure to effectively harness parallelism in the computation and efficiently utilize system resources. Our research designs, analyzes and implements scheduling strategies that are measurably good and practically efficient to provide various quality-of-service guarantees on cloud service latency.

Andrew Sohn
Associate Professor
sohna@njit.edu

Research Area: GPU Cluster Programming, Linux Kernel Development

GPU Cluster Programming for Solving Large-Scale Generative AI Problems

Recent reports revealed that the ChatGPT-4 model, which consists of over 1.76 trillion parameters and was released in March 2023, utilized over 30,000 Nvidia A100 GPUs, each with well over 10,000 cores, resulting in a total of over 300 million GPU cores. Given the enormity of 1.76 trillion parameters and the trillions of samples used to train the model, programming such a vast number of machines, GPU cores and samples is undeniably challenging. Our research is dedicated to developing scalable and enabling technologies for addressing large-scale generative AI problems using MPI and CUDA. We have been experimenting with this technology on a small scale and are currently expanding to build a laboratory aimed at establishing a large-scale problem-solving platform with MPI and CUDA.performance of a cyber-physical system.

Predictive Analytics on a Cluster of Computers

Predicting viral events in social networks in real time is challenging as events can unfold in a matter of months, weeks, days or even hours. We have been developing a system of hardware and software for real-time analytics, specifically targeting the prediction of viral events in social networks using a cluster of computers. One of the key enabling technologies is inertial spectral graph partitioning, developed in collaboration with NASA and Berkeley Lab. We have successfully implemented the framework to predict viral events in social networks, achieving over 70% accuracy on large-scale dynamic temporal Wikipedia graphs. We are currently working on further improving prediction accuracy.

Computing Education Research

Computing education research (CER) forges new discoveries in the design and application of teaching and learning technologies related to computing, broadly defined, with a focus on how people can benefit pedagogically from using these interventions. CER is an example of discipline-based education research, which is part of the broader field of education and learning science research.

Pictured: Turing Laureate Jeffrey D. Ullman, Ph.D.

Michael J. Lee
Associate Professor
michael.lee@njit.edu

Research Area: Human-computer interaction (HCI), computing education research (CER)

Advancing STEM Education with Gaming and Virtual Reality

We explore unique approaches to teaching STEM topics in formal and informal learning environments. Our work on Gidget—an online game (helpgidget.org) to teach students introductory programming concepts—is effective in attracting and engaging a broad audience, including women and underrepresented minority groups in computing. Our work with CSpresso uses an interactive VR environment to teach middle school and high school students introductory computer science concepts such as binary counting and sorting algorithms.

Increasing Diversity in STEM Through Mentorship

We explore how to increase participation and diversity in computing, especially for underserved and underrepresented minorities in STEM. We partner with local nonprofits and schools to provide programming experience to K-12 students, specifically using near-peer mentors to teach and engage middle school and high school students in a programming camp called Newark Kids Code. We also provide an introductory computing course for in-service high school teachers during the summer.

Jamie Payton
Dean
jamie.payton@njit.edu

Research Area: Pervasive computing, mobile sensing, Human-computer interaction (HCI), computing education research (CER), broadening participation in computing (BPC)

Connecting data structures and algorithms assignments to dynamic, real world data set

The BRIDGES project aims to enable the creation of more engaging, socially relevant assignments in introductory computer science courses by providing students with a simplified API that allows them to populate their data structure implementations with live, real-world data sets, such as those from popular social networks and web repositories (e.g., Twitter, Facebook, IMDb). The BRIDGES software system also allows students to create and explore visualizations of each executed data structure that they implement, which can promote better understanding of data structures and underlying algorithms. Students using BRIDGES show increased learning gains and successful progression in following courses in the CS major compared to a control group. Ongoing projects explore the integration of AI into the BRIDGES software platform to support the identification of individual student knowledge gaps and generation of customized challenge problems that connect to their prior, demonstrated proficiencies.

Computing Education Research

Integrating K12 computer science with physical activity and core subjects through wearable computing

To address the need for all K12 students to develop knowledge about computer science concepts, Project moveSMART aims to integrate computer science learning objectives into core subject areas, while motivating physical activity to promote student wellness as part of a whole learner approach. Project moveSMART uses a web-based platform and wearable sensors to integrate opportunities for physical activity with computer science and computational thinking (CS/CT) learning activities. The associated Project moveSMART curriculum embodies a project-based learning approach to expose students to data analysis and machine learning in the framework of classification and pattern recognition.

Developing Evidence-based Approaches for Broadening Participation in Computing

We build upon theories of identity development and student engagement to construct evidence-based approaches that broaden participation of underrepresented groups in computing and improve learning outcomes for all students. Focus is given to exploring pedagogical methods for teaching computer science that increase persistence in computing degree programs. Approaches are realized and experimentally validated through the STARS Computing Corps (STARS), a national alliance of more than 70 colleges and universities with a shared commitment to broadening participation in computing. At each STARS member institution, faculty support college students in leading computing-related service learning projects for the K-12 population, such as engaging students from underserved schools in computer science learning experiences or providing teachers with support for inclusive, engaging computer science learning activities. These projects promote persistence and advancement among STARS college students by 1) connecting their computing knowledge to community and societal impacts; 2) reinforcing their computing knowledge through teaching and practice; 3) developing professional, entrepreneurial and creative skills; and 4) fostering a network of computing faculty and students with a shared commitment to diversity, equity and inclusion in computing.

Computer Graphics

Computer graphics focuses on algorithms, software and methods for digitally synthesizing, understanding and manipulating visual content, ranging from optimizing designs for 3D printing, geometric modeling, computer vision, spatial computation, visual effects for movies and games, to AI and machine learning methods for both real-world and visual applications. NJIT researchers work both on improving the state-of-the-art and on conducting groundbreaking research in the above areas, for example, by using deep learning to synthesize virtual worlds and their AI agent inhabitants. Our research impacts many industries, including the entertainment industry in gaming and film-making, education, workforce training, health, virtual reality, military and computer-aided design for manufacturing.

Craig Gotsman
Distinguished Professor
craig.gotsman@njit.edu

Research Area: Computer graphics, geometric modeling, geometry processing, computational geometry

Distance Computations on 3D Models and General Networks

Computing geodesic distances and paths on 3D manifolds is an important problem in geometry processing. Existing approximate algorithms may efficiently compute an entire "distance field" from a single point on the manifold, but computing arbitrary point-to-point geodesic distances efficiently is difficult. Our work takes advantage of an efficient (pre- processed) hierarchical subdivision of a discrete 3D mesh, when treated as a graph, based on compact and balanced cuts, to devise an efficient "divide and conquer" online algorithm to answer these queries very quickly. Our basic methodology is generalized in a number of ways: 1) To compute lower-bound heuristics for shortest-path distances on road networks, which enables efficient shortest-path computation with the A* algorithm. 2) To efficiently approximate resistance distances on social networks. 3) To perform compact message routing (using small routing tables) with modest stretch in sparse networks.

Przemyslaw Musialski
Associate Professor
przemyslaw.musialski@njit.edu

Research Area: Computer graphics, geometric modeling, geometry processing, computational fabrication

Geometric Modeling and Computational Design

The traditional separation of design and manufacturing is currently one of the bottlenecks in the product development process leading to multiple expensive feedback loops. Our research aims at developing computational tools that integrate expert knowledge in 3D geometric modeling, physical simulation and smart algorithms in order to facilitate this process, which will become an important factor for success in increasingly competitive digital markets.

Computational Fabrication

We perform research at the intersection of applied mathematics, computer graphics and computational design. We investigate problems in geometric modeling and geometry processing, shape analysis and synthesis, as well as physical modeling and simulation with the goal of developing novel algorithmic solutions for the rapidly-growing domain of digital fabrication. Our research aims to provide novel solutions for the future design and manufacturing market.

Tomer Weiss
Associate Professor
tweiss@njit.edu

Research Area: Machine learning, data driven, optimization, simulation, crowd dynamics, deep learning, reinforcement learning

Computational Design of Virtual and Real Worlds

Virtual worlds are growing in complexity and interactivity and are challenging to construct efficiently and realistically. Interior scenes are an intrinsic part of virtual worlds. They are also needed for non-interactive applications, such as furniture marketing, residential showcasing and other commercial applications. Unfortunately, most of the aesthetic and creative process of designing interiors, real or virtual, is manual. To accelerate this process, we utilize machine learning and computer vision to develop several computational interior design tools. Our research assists users by providing suggestions on which furniture to pick, how well it matches the room's current style and where to place it. Hence, we can facilitate content creation needs for multiple industries and, most importantly, accommodate our common human needs to create our own optimal interior environment.

Simulating the Motion of Multi-Agent Crowds

Crowd simulation is an essential part of multiple domains, from interactive media systems, to pedestrian analysis in evacuation scenarios and urban planning, to autonomous multi- agent path planning and coordination. A crowd is a collection of independent, self-actuated agents, where each agent has individual navigational goals in this shared environment. Agent movement is controlled by a navigation algorithm, which needs to ensure that an agent progresses towards its goal while avoiding collisions. Computing collision-free agent motion is difficult, due to the complexity of such dynamic interactions. Despite more than 30 years in crowd simulation research, simulation methods have subspecialized and are computationally effective for either sparse or dense crowds but not for both. Our work is the first method that allows real-time simulation of both dense and sparse crowds for hundreds of thousands of agents. This result is made possible by reframing agent motion as a constrained mathematical optimization problem that can be solved effectively on modern GPU computer hardware. Aside from the immediate practical implications for the gaming and visual effects industry, these results inspire future research on large-scale pedestrian dynamics which can affect urban and space planning.

Cybersecurity

The security research group performs extensive work in areas such as software supply chain security, web security and privacy, smartphone and mobile security, database security, trustworthy ML/AI, homomorphic encryption and human-centered security and privacy. It also makes significant contributions to open-source software. The group's research has been externally funded by agencies such the NSF, DARPA and the NSA and members are recipients of two NSF CAREER awards and one DARPA Young Faculty Award. The group aims to publish its research in top security and related venues, such as Usenix Security Symposium, IEEE Security and Privacy Symposium, ACM Conference of Computer and Communications Security, ACM Mobicom and ACM SIGMOD.

Reza Curtmola
Professor
reza.curtmola@njit.edu

Research Area: Cybersecurity, software security, web security and privacy

Defending Software Supply Chains Against Hackers

Funded by DARPA and the NSF, we have developed in-toto, an open-source framework that promises to safeguard software for developers and end users. In-toto provides organizations with insights into the software development and distribution chain, such as having a provable assurance that proper software development practices were followed. With in-toto in place, it will be more difficult for malicious code to be slipped into software products, thus raising the bar significantly for attackers. Through integrations, in-toto is currently used by thousands of companies and has improved the security of millions of users.

Web Security and Privacy

The goal of this project is to explore targeted privacy attacks on the web through the lens of side channels. We uncovered new attacks that can lead to targeted deanonymization on the web by using CPU cache side channels. In particular, we uncover a set of practical and scalable attacks that can deanonymize users in several important settings for which prior attack methods are not effective. This affects all major browsers, including Chrome, Firefox, Safari, Edge, Tor Browser and numerous major sites, including Google, Twitter, LinkedIn, TikTok, Facebook, Instagram and Reddit. Our attacks run in less than 3 seconds in most cases and can be scaled to target an exponentially large number of users. More importantly, we provide a comprehensive countermeasure against all of the attacks we discovered. This countermeasure is already available on the Chrome and Firefox extension stores and can be downloaded and installed immediately by concerned users. This work was published in the 31st USENIX Security Symposium.

Iulian Neamtiu
Professor
iulian.neamtiu@njit.edu

Research Area: Programming languages, software engineering and their applications to reliable AI, smartphones, security

Android Security

Our research is focused on security issues in Android apps and the Android platform, including exposing deceptive practices in apps, apps attempting to cover their traces, ransomware, unauthorized collection and transmission of user data in general and personally identifiable health information in particular, apps refusing to disclose the data they collect or refusing to delete data when legally mandated.

Shantanu Sharma
Assistant Professor
shantanu.sharma@njit.edu

Research Area: Database, security, privacy, blockchain, IoT

Information-Theoretically Secure Processing

Despite over two decades of research, secure data outsourcing remains an open challenge. Information-theoretically secure techniques provide the highest level of security regardless of the computational capabilities of an adversary. One of the well-known information-theoretically secure techniques is Shamir's secret sharing. We develop information-theoretically secure data processing systems that can efficiently execute different types of SQL queries on large databases. Furthermore, we focus on information-theoretically secure machine learning techniques.

Smart and Privacy-Preserving Smart Spaces

Smart spaces are rapidly growing in present time. Examples of smart spaces are office/university buildings, shopping malls, train/bus stations and airports that capture user-related data via different types of sensors. While such sensor data is beneficial to developing multiple value- added services, smart spaces jeopardize user privacy due to mixing sensor data with the digital representation of space. For example, tracking a person in real-time can reveal their behavior. We develop an end-to-end secure and privacy-preserving smart space that respects user privacy at each stage of data processing, such as data collection, storage, processing, sharing and auditing.

Cong Shi
Assistant Professor
cs638@njit.edu

Research Area: Mobile Security

Privacy Study of Unrestricted Motion-Position Sensors in the Age of Extended Reality

Extended Reality (XR) has gained popularity in numerous fields. We conduct a comprehensive study to assess the trustworthiness of the embedded sensors on XR, which embed various forms of sensitive data that may put users' privacy at risk. We find that accessing most on-board sensors (e.g., motion, position and button sensors) on XR SDKs/APIs requires no user permissions, exposing a huge attack surface for an adversary to steal users' private information, such as keystrokes, speeches and sensitive physiological states.

Zhihao "Zephyr" Yao
Assistant Professor
zy8@njit.edu

Research Area: Operating systems, mobile computing, security and privacy, system support for secure AI/ML

Strengthening Trust and Security in Mobile Systems

In our increasingly digitized society, ensuring the security of mobile systems has become critical, especially as smartphone owners often run security-critical financial applications and life-critical medical applications alongside untrusted programs. Our research focuses on creating novel solutions to strengthen trust in these devices, particularly in scenarios where security and privacy are challenged by the increasingly complex software stack and the integration of artificial intelligence. By minimizing the Trusted Computing Base (TCB), designing trusted execution environments and developing hardware-based isolation techniques, we aim to provide robust safeguards against privacy leakage, unauthorized access and service disruption, while maintaining performance and usability. These efforts enable secure and reliable mobile systems that support a range of novel applications, from verifying video authenticity to ensuring the reliability of high-assurance medical systems.

Enhancing Privacy and Security in Large Language Model Interactions

As Large Language Models (LLMs) become more integrated into online experiences, the privacy and security implications of their use are increasingly concerning. Our research aims to address these challenges by developing methods to protect user privacy and enhance the security of LLM interactions. We focus on mitigating privacy risks by sanitizing sensitive user data before it reaches LLM services and examining the security flaws in LLM-generated code. Additionally, we explore the security implications of the emerging WebGPU interface, which is increasingly used in conjunction with LLMs to accelerate performance in a web browser. Our goal is to make the interactions with LLMs more secure for everyday users.

Nathan Malkin
Assistant Professor
nathan.malkin@njit.edu

Research Area: Human-centered security and privacy

Human-Centered Security and Privacy

Our research investigates how human factors — people's distinctive priorities, abilities and limitations, as well as group behaviors and processes — contribute to cybersecurity and privacy failures and how we can improve our systems to avoid these problems. For example, we have studied how system administrators deploy software updates, developers write secure code and end users manage their privacy with respect to voice assistants. We then design and empirically validate systems that help people overcome these technological challenges through more usable interfaces, such as by automating and eliminating confusing choices in smart home settings.

Data Mining

The abundance of available data, such as images, time series, trees and graphs, allows for educators and governments to gain previously unavailable insights through a deeper understanding of the data and make better predictions with real-world datasets. Our group is focused on discovering knowledge from data and developing tools that improve the process of data analytics. These analytics have created real-world, trustworthy intelligence. We work on a variety of techniques, including pattern mining, deep learning and generative models. These techniques provide tools to address issues ranging from data summary and search, to time series analysis, to space weather forecasting.

Jason Wang
Professor
wangj@njit.edu

Research Area: Data mining, machine learning, deep learning, explainable AI, generative AI, trustworthy AI, data science

Mining Big Data Through Deep Learning

We are designing new deep learning algorithms for mining big data. We have developed a 3D-atrous convolutional neural network, used it as a deep visual feature extractor and stacked convolutional long short-term memory networks on top of the feature extractor. This allows us to capture not only deep spatial information but also long-term temporal information in the data. In addition, we use stacked denoising autoencoders to learn latent representations of the data, which can be used to construct feature vectors suitable for classification. We also develop new recurrent neural networks to mine time-series data for space weather prediction. Currently, we are building a deep learning framework with generative adversarial networks and conditional diffusion models. The framework can handle model uncertainty as well as data uncertainty and sparsity. Our deep learning models are suited for big data applications that have few, incomplete, imperfect, missing, noisy or uncertain training data.

Dimitri Theodoratos
Associate Professor
dimitri.theodoratos@njit.edu

Research Area: Data mining, pattern extraction

Mining Big Data Through Deep Learning

Extracting frequent patterns hidden in trees and graphs is critical for analyzing data and a first step for downstream data mining. Most pattern-mining algorithms do not scale to big data applications. We have designed algorithms to extract patterns from large trees and graphs, leveraging results using compressed bitmap views.

Searching Structured and Semi-Structured Data with Keyword Queries

Disambiguating a user's intention in posing a keyword query and efficiently retrieving relevant results is an immense challenge for keyword search when using big data. We have devised an approach that exploits a structural summary of the data to extract pattern graphs for keyword queries. This empowers non-expert users to extract information from data sources and databases without mastering a query language and without any knowledge of the organization or structure of data sources.

Extended Reality

Our research is interdisciplinary and covers the spectrum of virtual reality, augmented reality and mixed reality for a wide range of immersive applications where elements from gaming, 3D graphics, digital media and simulations are used. We design and develop innovative applications to enhance learning, cognition, communication, decision-making, usability and user experience in fields such as health care, education, the military, manufacturing and science, pushing the boundaries of immersive technology. Examples of projects include a Brain-computer augmented reality environment to support and trace complex consumer decision-making, political decision-making and psycho-therapeutic applications; virtual geriatric patients that aim to improve caregiver training through interactive simulations; anywhere simulation software that facilitates remote healthcare education; advanced visualization techniques for large datasets and immersive simulations, forensic research and digital histories; and measuring and tracking wound progressions in 3D on complex surfaces.

Frank Biocca
Professor
frank.biocca@njit.edu

Research Area: Virtual and augmented reality systems, components for brain-computer interfaces, real-time public opinion measurement

Design of Virtual Environments and Interfaces to Support Information, Perception and Cognition

Our research focuses on designing VR and AR hardware and software to enhance user cognition and performance across medical, scientific and military applications within the Media Interface and Network Design labs. In collaboration with teams in Spain and Korea, we explore how a brain-computer AR environment can aid complex decision-making and therapeutic processes, using untethered brain and psychophysiological sensors to detect how virtual features influence thinking and enabling real-time environment adaptation. Other projects investigate how VR environments alter body perception and social cognition in negotiation, training and decision-making. In scientific visualization, our AR environments allow users to experience physical forces or microscopic phenomena, as in an astrophysics project where magnetic sensor data visualizes Earth's local magnetic fields in real time.few, incomplete, imperfect, missing, noisy or uncertain training data.

Jacob Chakareski
Associate Professor
Jacob.chakareski@njit.edu

Research Area: Immersive communication, augmented/virtual reality

Virtual Human Teleportation

Virtual reality and 360-degree video are emerging technologies that can enable virtual human teleportation to any remote corner of the globe. This requires ultra-low latency, gigabit-per-second wireless speeds and data-intensive computing. Our research investigates synergies at the intersection of 6DOF 360-degree video representation methods, edge computing, UAV-IoT, millimeter-wave and free-space optics wireless technologies. It transmits data using much higher electromagnetic wave frequencies to enable the ultra-high data rates and ultra-low latencies required by next generation societal VR applications.

Real-Time Structure-Aware Reinforcement Learning

Reinforcement Learning (RL) provides a natural paradigm for decision-making in diverse emerging applications that operate in unknown environments and with limited data of unknown stochastic characteristics. Paramount to the effective operation of these ultralow latency applications, such as IoT sensing, autonomous navigation and mobile virtual and augmented reality, is the ability to learn the optimal operation actions online and as quickly as possible. Existing state-of-the-art RL methods either take too long to converge or are too complex to deploy. Our research examines novel structure-aware RL methods that integrate basic system knowledge to compute learning actions updates across multiple states or even the entire state-space of the problem of interest, in parallel. To address the challenge of computational complexity that is introduced, our methods integrate analysis that help effectively trade-off learning acceleration and computing complexity.

Societal Applications

Our research focuses on interdisciplinary synergies to enable next-generation applications. For instance, a National Institutes of Health project at the intersection of networked virtual reality, artificial intelligence and low-vision rehabilitation aims to enable novel, previously inaccessible and unaffordable health care services to be delivered broadly and affordably. Other projects include the integration of virtual reality, real-time reinforcement learning and soft-exoskeletons for future physical therapy and the synergy of UAV-IoT and VR towards next generation forest fire monitoring.

Salam Daher
Assistant Professor
salam.daher@njit.edu

Research Area: Augmented, virtual and mixed reality; physical-virtual, 3D graphics; virtual humans; synthetic reality; modeling simulation and training; distance simulation; healthcare applications and virtual patients; 3D wounds visualization, measurement and tracking

Mixed Reality Simulations to Improve Training

Our research focuses on creating simulations using computer graphics, multimedia and mixed reality to improve training in different domains including health care simulation. We are especially interested in research involving virtual humans and multisensory experiences. We developed a new class of augmented reality patient simulators called physical-virtual patients that allow health care educators to interact with a life-size simulated patient by providing real-time physical tactile cues such as temperature and pulse; auditory cues such as speech and heart sounds; rich dynamic visual cues such as facial expressions indicating pain or emotions; and changes in appearance such as skin color and wounds.

Training Caregivers of Virtual Geriatric Patients

We are developing a new generation of Virtual Geriatric Patients (VGP). The VGPs are realistic, embodied, conversational virtual humans who are aware of their surroundings. The VGPs are displayed in Mixed Reality as training scenarios aimed to improve caregivers' perceptions, attitudes, communication and care towards older adults. This research is supported by a grant from the National Science Foundation Future of Work at the Human Technology Frontier.

3D Graphics for Wound Visualization, Measurements and Tracking

Our research focuses on visualizing wounds in 3D for accurate measurements, reduced variability of measurements and improved tracking of patients' progress. In the clinical setting, this translational research can reduce errors, improve healing estimates and improve patient outcomes. In the training setting, this technology can improve healthcare trainees' skills in wound assessment, especially when combined with mixed reality. This research is partially supported by the New Jersey Health Foundation and by the NJIT Technology Innovation, Translation and Acceleration (TITA) seed grant.

Interactive Remote Simulation for Healthcare Training

During the pandemic, healthcare educators rushed to use pre-existing videos or had to record their own videos that they shared with their students to watch as a makeshift "simulation". Content needs to be interactive to satisfy the interactivity requirement for simulation. Our team developed software called "Anywhere Simulation (AwSIM)" that allows healthcare educators to add interactivity to their existing content to create new healthcare scenarios and share that content remotely with their students. AwSIM provides healthcare educators with the power to create their own simulation scenarios using their own content (e.g., videos, images, text) and run the simulation remotely with their students. The software is content independent and is easy to use. We ran multiple studies with nursing students using our AwSIM technology and found that adding interactivity promotes teamwork, perception of authenticity and higher levels of thinking. Also, the AwSIM software has a high technology acceptance rate among students. We are working on creating and evaluating an immersive standalone version for trainees that they can use on a flat screen or on with a head mounted display with or without a facilitator.

Erin J.K. Truesdell
Assistant Professor
ejt25@njit.edu

Research Area: Human-computer interaction, games, game design, serious games, extended reality, tangible interfaces, collaboration, play

Design for Multi-User XR Experiences

Recent technological developments have made it easier than ever to design games, tools and even educational experiences in virtual and augmented reality. Tetherless headsets, hand tracking tools and high-quality video passthrough allow for a vast new world of interactive media to be created. This project aims to better understand how these technologies can be leveraged to create engaging and informational experiences where users share physical space as well as digital space. Specific research topics include group navigation strategies, novel input systems for games and space layouts, interaction design and audio considerations for setups with multiple co-located users.

Margarita Vinnikov
Assistant Professor
vinnikov@njit.edu

Research Area: Immersive and collaborative extended reality (AR/VR/MR), navigation, gaze/body tracking, simulations

Visualization of Large Datasets & Collaborative Platforms (VROOM Project)

Our VROOM project, in collaboration with Professor James Geller's SABOC (the Structural Analysis of Biomedical Ontologies Center), focuses on developing advanced methods for visualizing large datasets in virtual reality environments. The project is particularly concerned with the challenges of rendering complex datasets, such as ontology trees and visibility graphs, within XR systems. These visualizations are integrated into multi-user, multi-player platforms where users can collaboratively interact with the data in real time. Our research combines cutting-edge user-centered design principles to enhance the efficiency and usability of large-scale data visualization in immersive environments.

Forensic Research (CSIxR)

We explore how XR technologies can be applied in forensic science to revolutionize crime scene analysis and forensic training. Our work involves creating detailed, immersive virtual environments that simulate crime scenes, enabling forensic professionals to analyze evidence and reconstruct events with a high degree of precision. These simulations are also used for training purposes, providing an interactive platform to practice forensic techniques and methodologies. By integrating XR into forensic research, we aim to improve both the accuracy of investigations and the efficiency of forensic processes.

Digital Histories

In collaboration with Dean Louis Hamilton and Associate Dean Burcak Ozludil (Albert Dorman Honor's College), the Digital Histories project focuses on creating immersive historical narratives through the use of virtual reality and digital databases. This includes the development of digital catalogs and visual archives that offer rich, interactive experiences for exploring historical events and sites. By leveraging multimedia, online data and VR visualizations, we provide users with a deeper understanding of history through engaging and immersive platforms.

Immersive Cross-Reality Applications

Our research in immersive cross-reality applications spans a wide range of projects, including augmented reality platforms and VR simulations. This includes the development of an augmented reality chemistry platform, industrial training tools and HUDs for airplane parking. Additionally, we create VR simulations for driving, flying and exploring animal habitats. These projects focus on creating highly interactive, multi-sensory experiences that push the boundaries of what XR technologies can achieve in both educational and industrial settings.

Human Computer Interaction

Human-computer interaction (HCI) is a multidisciplinary field of study that focuses on how people interact with technology. This technology can range from physical artifacts such as sensors and wearable devices, to networked systems such as social media and online communities and algorithms (such as AI and generative AI). HCI can be applied to almost all aspects of society, in domains such as computer-supported collaborative work, health, education, security/privacy, music, cyberpsychology, social computing and entertainment using qualitative, quantitative and data visualization methods.. The field also includes the study of technology use with special populations to examine accessibility and inclusive technology. This work is frequently published in top-tier venues such as ACM CHI, ACM CSCW, ACM ASSETS, TVCG and IEEE VIS.

Julie Ancis
Distinguished Professor
jra49@njit.edu

Research Area: Cyberpsychology, human-computer interaction, tailored social media messaging, misinformation, online hate, social media infleuncers, disinformation, bias

Cyberpsychological Impacts of Online Information

The informational landscape and how individuals and groups engage with this information has implications for a broad range of human behaviors. Using qualitative and quantitative approaches grounded in psychology and computing, we analyze how both the text and visual components of social media messages are constructed to convey information about issues ranging from health crises to critical societal events. Our research investigates the impact of these messages on risk perception, decision making, proactive health behaviors, civil discourse and bias. This work has four major foci: 1) The evaluation of tailored and targeted online messaging. 2) Development of culturally informed online information and messages to enable adaptive behavior. 3) Testing the psychological and behavioral impact of culturally relevant messaging on perceptions and intentions.

Mark Cartwright
Assistant Professor
mark.cartwright@njit.edu

Research Area: Machine listening, interactive machine learning, human-computer interaction, audio processing, music information retrieval, sound accessibility

Tools for Accessible Sound Understanding and Creation

Sound is all around us and in the media we consume. However, for the roughly 1 in 5 people who are to some degree deaf or hard of hearing (DHH), the information contained in this medium is not always accessible. In this project, we aim both to develop tools to increase the accessibility of information contained in sound and to develop tools to make the creation of audio artifacts more accessible.

Aritra Dasgupta
Assistant Professor
aritra.dasgupta@njit.edu

Research Area: human-AI teaming, interactive visualization, visual analytics, responsible AI, human-machine communication

Human-AI Collaboration using Visual Analytics

Our research group aims to address the question: "How can we bridge the reasoning gap between people & data-driven models?" As newer technologies such as large language models and generative AI continue to be developed and adopted in socio-technical applications, it will be increasingly critical to address this reasoning gap so that human stakeholders have appropriate levels of agency in interpreting, verifying and partaking in the model-backed claims and decision-making processes. Complex algorithms and AI models often drive decisions based on available data. Be it climate scientists simulating global warming, epidemiologists forecasting regional pandemic trajectories, or digital interfaces recommending products to consumers, data needs to be modeled reliably to provide real analytical value. However, developing and interpreting data-driven models, like those resulting from machine learning or scientific simulations, necessitate substantial human effort to understand the context and significance of patterns in the data.

To address these challenges, we conceptualize, design and evaluate visualization as an evidence-based communication medium between humans and complex AI models. The techniques we build integrate automated methods such as statistics and machine learning with optimal visualizations to enable human judgment and reasoning about AI-detected patterns.

This integration ensures that we leverage the best of both worlds: computational power for fast extraction of patterns (e.g., a predictive model trained on millions of transactions for flagging suspicious individuals) and our perceptual and cognitive faculties for letting experts transparently reason about the context and significance of the patterns (e.g., financial analysts visualizing past and present behavior of individuals to understand the context of predictions). The artifacts of our research are evaluated using a mix of quantitative (e.g., accuracy metric) and qualitative (e.g., design study and case study) methods. My research methodology entails participatory design of visualizations and interactive systems, empirical evaluation of techniques through user studies and observational techniques to study how people interact with data and AI models.

Amy Hoover
Assistant Professor
amy.k.hoover@njit.edu

Research Area: Optimization

Methods in Quality Diversity

While many optimization techniques in evolutionary computation maximize fitness with respect to one or more objectives, often such searches restrict the set of candidate solutions to those with objective values lying along the Pareto front of optimality. Instead, algorithms in Quality Diversity (QD) fully explore these objectives by specifying them explicitly as dimensions (called behavioral characteristics) that are characterized by their genomic, phenotypic or behavioral traits. Such exploration can not only generate a large collection of high-performing solutions, but with well-chosen dimensions can potentially find higher performing solutions than pure-objective based searches alone.

Sooyeon Lee
Assistant Professor
sooyeon.lee@njit.edu

Research Area: Human-computer interaction, accessibility, human-AI interaction

Design and Evaluation of Accessible AI Technologies for Users with Disabilities

Over one billion people in the world live with some type of disability. Many of them experience barriers in accessing information or using technologies, which can limit social interactions in both physical and digital spaces. Our work investigates the diversity of users, explores and leverages emerging technologies and adopts human-centered AI and inclusive design approaches in the design and evaluation of new AI based systems and applications that address accessibility barriers. Our research focuses on investigating and designing non-visual interaction for the community of blind users and non-audio and non-speech interaction for the community of deaf and hard of hearing users.

Alisha Pradhan
Assistant Professor
ap2972@njit.edu

Research Area: Human-computer interaction, human-AI interaction, inclusive design, aging, accessibility, underrepresented populations

Towards Inclusive Design of Emerging Technologies

In today's world, our dependence on technology has become so profound that it influences nearly every aspect of our daily lives. But when it comes to design of such technologies, who do product designers typically consider as users? More importantly, who do they exclude as users (even though, inadvertently)? And what barriers and challenges do those users, who are rarely considered in the technology design process, experience when using these systems? What harmful assumptions or stereotypes do our technologies encode about them? These are some of the questions that guide our research on inclusive design of emerging technologies. Older adults represent one group, typically underrepresented in technology design, where their perspectives and preferences are rarely included. By using user-centered design approaches, we closely work with older individuals in design and development of technologies. We examine the benefits and barriers posed by existing technologies, as well as adopt participatory approaches to design technologies for and with older individuals.

Erin J.K. Truesdell
Assistant Professor
ejt25@njit.edu

Research Area: Human-computer interaction, games, game design, serious games, extended reality, tangible interfaces, collaboration, play

Understanding Collaboration with Tangible Interfaces

Our research centers on how to design novel tangible interfaces to support multi-user collaboration with a particular focus on playful and creative applications. We are interested in understanding how people work together using a variety of tangible and wearable interfaces to achieve goals, produce creative work, develop interpersonal skills and relationships and learn new information through interaction. Insights gained from this research have applications in entertainment and beyond, in fields such as computer-supported collaborative work, formal and informal learning and human-AI interaction.

Yvette Wohn
Associate Professor
wohn@njit.edu

Research Area: Human-computer interaction, social media, digital economy, online self-presentation

Social Media and Wellbeing

This research investigates the relationship between psychological well-being and technology usage, especially how social media can play a role in facilitating both hate and social support. This includes the study of individualistic behavior—such as understanding how people form and present their identity online—as well as collective behavior, such as how people engage in collective action and activism.

New Digital Economies

The metaverse may seem like a new phenomenon but research on virtual spaces has been going on for decades. In this research, we focus on systems with novel digital economies in virtual environments that have unique digital currency, such as online games. Our research examines spending behaviors and exchange patterns of virtual goods in games and other alternative financial platforms and how these activities are tied with creative content generation.

Image and Video

Our research on image and video processing continues to advance science and technology. We work with experts in health care, state and local government, security and the private sector to examine obstacles and address them by leveraging the latest advancements. From the diagnosis of pneumonia and brain tumors to automated systems that can monitor traffic incidents, roadway congestion and public safety risks, this group is making advances with significant work in image forensics and watermarking, pattern and face recognition, contrastive learning, image captioning, medical image analysis, video analytics, deep neural networks and smart transportation systems.

Chengjun Liu
Professor
chengjun.liu@njit.edu

Research Area: Computer vision, pattern and face recognition, video processing

Making Intelligent Transportation Systems Smarter

The New Jersey Department of Transportation designated more than 400 closed-circuit video cameras statewide for incident monitoring, traffic congestion control and public safety operations. Video streams from these cameras feed to a back-end system. There, video analytics software is used to perform target detection and incident monitoring applications. We are actively working on incorporating wireless sensor networks, hierarchical edge-computing and computer vision to mitigate the challenging problems in various illumination and weather conditions in order to achieve fast and automated video based traffic monitoring.

Video Analytics Pilot Studies and Testing of Technologies

We propose a new modular approach for statistical modeling of traffic incidents and model selection in order to improve state-of-the-art traffic detection and monitoring. We investigated and developed automated video analytics systems to replace human operators for traffic incident detection and to monitor the cameras installed along the major New Jersey highways. We test the proposed technologies and benchmark their performance.

Frank Shih
Professor
shih@njit.edu

Research Area: Image processing, artificial intelligence, computer vision, digital watermarking, digital forensics, robot sensing, neural networks

Deep Learning for Pneumonia Classification and Segmentation on Medical Images

Automatic identification of pneumonia on medical images has attracted intensive study. In this project, we develop a novel joint-task architecture that can learn pneumonia classification and segmentation simultaneously. Two modules, including an image preprocessing module and an attention module, are developed to improve both classification and segmentation accuracies. Experimental results performed on the massive dataset of the Radiology Society of North America have confirmed its superiority over other existing methods.

Deep Morphological Neural Networks and Applications

Given a target image, determining suitable morphological operations and structuring elements is a cumbersome and time-consuming task. In this project, we propose new morphological neural networks, which includes a nonlinear feature extraction layer to learn the structuring element correctly and an adaptive layer to automatically select appropriate morphological operations. We also use them for their classification applications, including hand-written digits, geometric shapes, traffic signs and brain tumors. Experimental results show higher computational efficiency and higher accuracy when compared with existing convolutional neural network models.

Adaptive Data Transformation in Contrastive Learning

Although contrastive learning has been playing a critical role in pattern recognition, how to optimize positive pairs through data transformation is still not well developed up to now. In this project, we propose a novel Adaptive Data Transformation, named ADTrans, to identify an optimal sequence of data transformations, which enables generating high-quality positive pairs adaptively during contrastive training. Extensive experiments on benchmark datasets have shown that ADTrans can improve the performance of representation learning on downstream tasks significantly, including image classification, instance segmentation and object detection.

Image and Video

Image Captioning with Contextual Feature Exploitation

Existing methods for image captioning lack the incorporation of sufficient contextual information in the generated captions. In this project, we present ContExCap, a novel and effective approach that exploits contextual features from an image to enhance its caption with more contextual information. Moreover, we introduce a new module, named Object-Context Attention, to capture the interaction and relationship between objects and contextual features. To utilize the low-level features, we also incorporate feature fusion of every second encoding layers with spatial shift operation that enables the features to align with neighbors.

Guiling "Grace" Wang
Distinguished Professor
gwang@njit.edu

Research Area: Deep Learning, AI in Finance, AI in Transportation, Large Language Model (LLM)

Reverse Pass-through VR and Full head Avatars

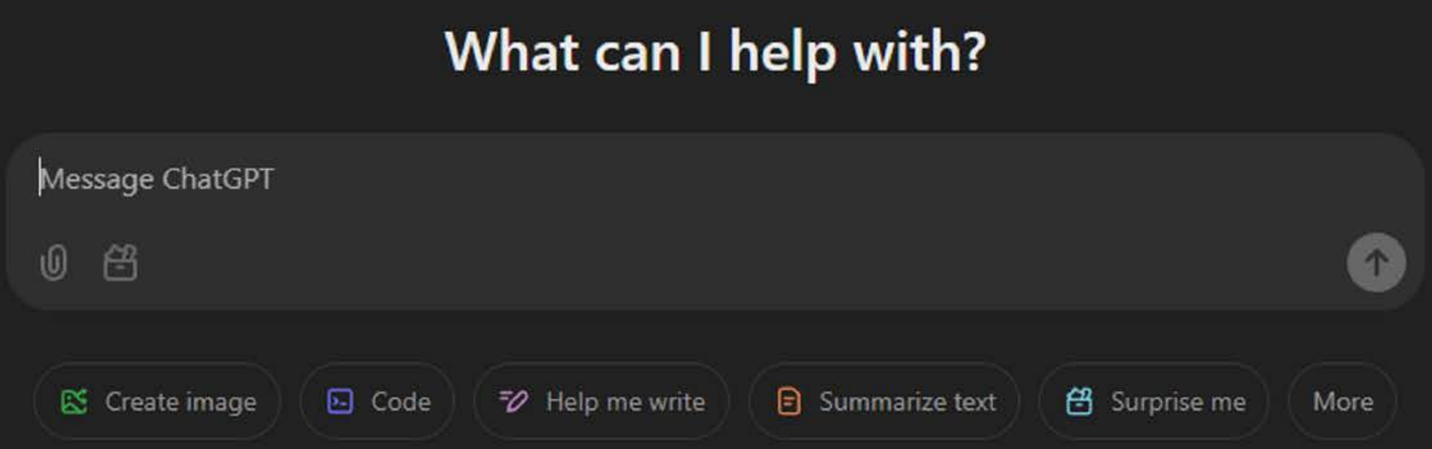
We introduce a novel, state-of-the-art method for reverse pass-through VR and one-shot full-head avatar generation. Our approach enables real-time reconstruction of complete facial images from partial VR inputs, delivering a seamless reverse pass-through experience. Using deep learning and 3DMM models, we generate photorealistic, one-shot full-head avatars from just a single image. The system functions in an end-to-end manner, ensuring an immersive VR experience. To support further research, we developed a new dataset to simulate complex VR conditions, including diverse occlusions and lighting variations. Our framework surpasses existing state-of-the-art methods in both visual quality and realism.

HierGAN: Combined RGB and Depth inpainting

We present Hierarchical Inpainting GAN (HierGAN), a real-time, novel approach for RGBD inpainting designed to overcome the limitations of current methods in Diminished Reality (DR) applications. HierGAN employs a hierarchical GAN architecture that leverages RGB, depth, edge and segmentation label images to effectively use RGB and depth in painting. By incorporating segmentation label images as auxiliary inputs, HierGAN significantly enhances inpainting performance, achieving superior results over existing methods with an end-to-end, fully optimized workflow.

Large Language Models

Our research on large language models (LLMs) focuses on advancing the capabilities, reliability and applications of these powerful AI systems. We work on developing more trustworthy LLMs by improving their explainability, fairness and robustness. Our efforts include techniques to uncover LLM functioning, enhance fairness in in-context learning and mitigate issues like shortcut learning and biases. We also investigate the theoretical foundations of LLMs, analyzing their training dynamics and emerging capabilities like in-context learning. Our work extends to practical applications, exploring LLMs' potential in mathematical reasoning, creative problem-solving and secure data analysis. From enhancing LLMs' ability to solve complex math problems to developing frameworks for DataFrame question answering without data exposure, we aim to push the boundaries of what LLMs can achieve while ensuring their responsible and effective deployment in real-world scenarios.



Mengnan Du
Assistant Professor
md748@njit.edu

Research Area: Trustworthy AI, explainability, fairness, natural language processing, large language models

Alignment of Multi-modal LLMs

Our research focuses on addressing the critical alignment issue in multi-modal Large Language Models (LLMs), with particular emphasis on Vision-Language Models (VLMs) and Protein-Language Models (PLMs). We investigate fundamental challenges in achieving robust alignment between diverse data representations through three key perspectives: visual-linguistic alignment, protein-structural alignment and cross-modal integration. Our work includes developing methods to bridge representation gaps between modalities, understanding alignment mechanisms in different architectures and improving model reliability through better training and evaluation approaches. By innovating in alignment techniques, evaluation frameworks and theoretical foundations, we aim to enhance the capability of LLMs to work seamlessly across different modalities while maintaining semantic consistency and factual accuracy.

Shuai Zhang
Assistant Professor
sz457@njit.edu

Research Area: LLMs, In-context Learning, Transformer, Mamba, MOE, Feature learning, Generalization guarantees

Advancing Leaning Theory with Emerging LLM Structural Components

LLMs have demonstrated exceptional performance across many domains. This success is largely attributed to their "emergent" abilities, enabling them to solve complex, unseen tasks not explicitly anticipated during pre-training. A key example is in-context learning (ICL), where LLMs learn new tasks from just a few contextual examples, without fine-tuning. These capabilities are believed to stem from transformer-based pre-trained models, though the theoretical mechanisms remain unclear. Our goal is to advance the current feature-learning framework by analyzing the training dynamics of transformer-based LLMs and investigating the unique role of the attention mechanism. Additionally, we aim to extend this framework to incorporate emerging architectural components in LLMs, such as structured state space models and Mixture-of-Experts and work towards designing more efficient model architectures with strong generalization guarantees.

Guiling "Grace" Wang
Distinguished Professor
gwang@njit.edu

Research Area: Deep Learning, AI in Finance, AI in Transportation, Large Language Model (LLM)

Enhancing Controllability and Explainability in Flowchart Understanding with LLMs

Flowchart understanding, often reliant on vision-language models (VLMs), faces challenges in controllability and explainability. Users have limited ability to influence processing beyond input modification and errors are difficult to trace due to opaque reasoning processes. To address these issues, we propose a two-stage framework: a VLM converts flowchart images into customizable text representations and an LLM performs reasoning and question-answering on the text. This approach enhances user control, isolates processing errors for improved explainability and promotes modularity by enabling integration with advanced reasoning tools. The framework's structured intermediate representations also provide a foundation for generalizing to other multimodal tasks, improving usability and reasoning capabilities.

Advancing Creative Problem-Solving in Mathematics with LLMs

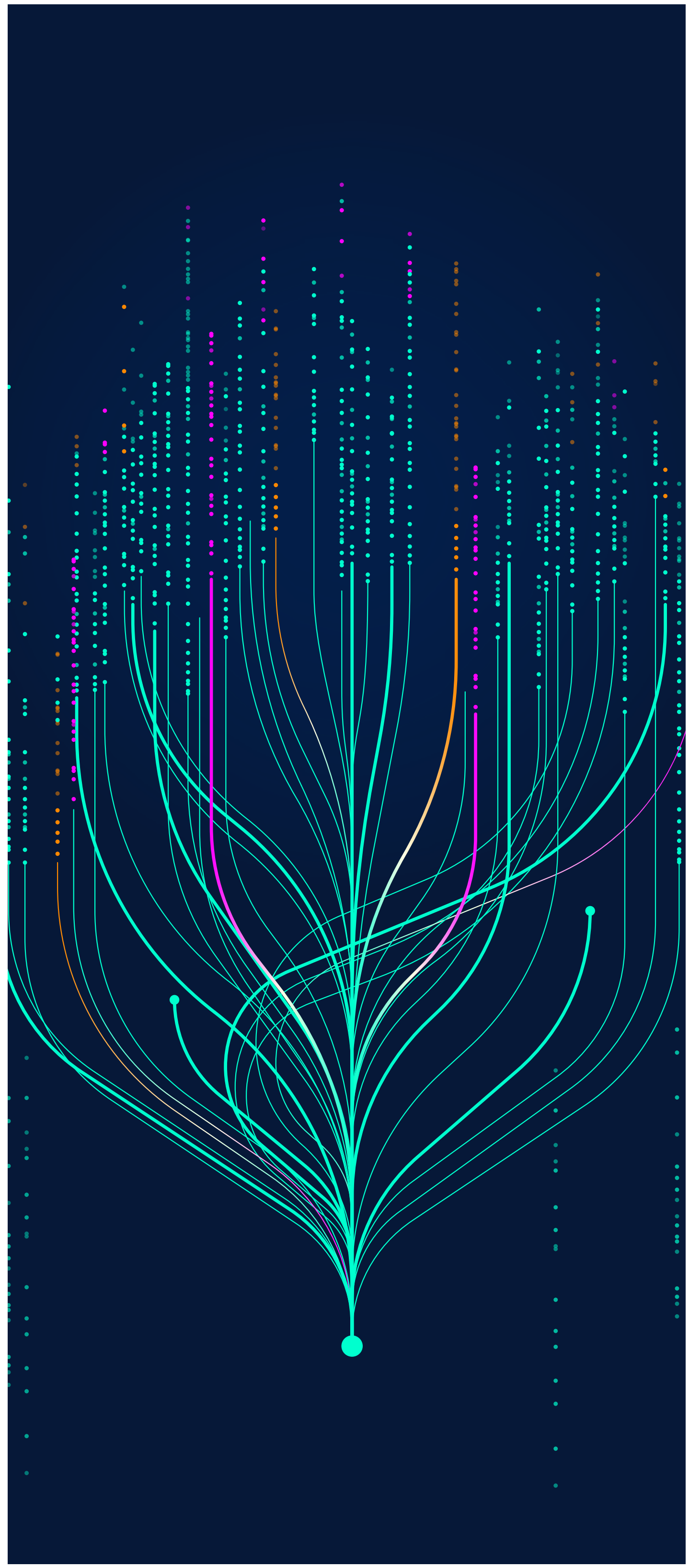
While research on Large Language Models (LLMs) has extensively explored their problem-solving capabilities, their potential for creativity in mathematical reasoning remains underexamined. To bridge this gap, we present CreativeMath, a framework designed to evaluate and enhance LLMs' innovative reasoning abilities in mathematics. Published at AAAI 2025, CreativeMath introduces a benchmark comprising problems spanning middle school curricula to Olympic-level challenges, systematically assessing the creative problem-solving skills of LLMs. This study sheds light on both the strengths and limitations of LLMs in fostering mathematical creativity and offers a robust benchmark to advance our understanding of their cognitive potential.

Advancing Logical Reasoning in LLMs with FaultyMath

Large Language Models (LLMs) excel at solving standard mathematical problems but often fail to detect logical inconsistencies, raising questions about their ability to reason beyond rote calculation. To address this, we present FaultyMath, a benchmark for evaluating LLMs' capacity to identify and reason about faulty math problems. FaultyMath encompasses diverse categories, including algebra and geometry, with varying difficulty levels and fault types such as contradictions and common-sense violations. It assesses LLMs' performance in detecting flawed problems, incorporating hints and providing reasoned explanations. This research underscores the limitations of current LLMs in logical reasoning and establishes a foundation for enhancing their cognitive capabilities, fostering more robust and trustworthy AI systems.

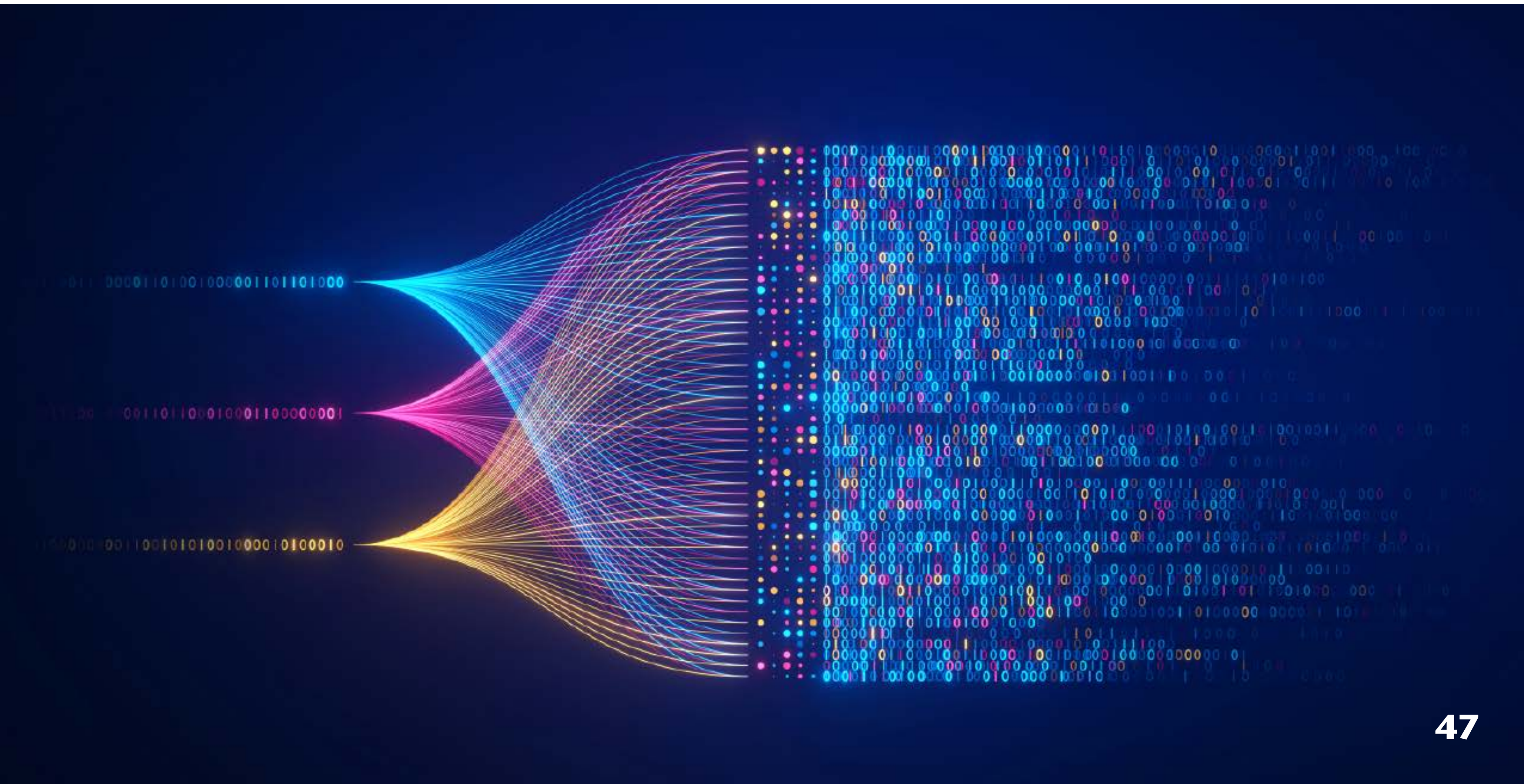
Revolutionizing Table Question Answering with Secure and Efficient LLM Solutions

Table-based question answering with Large Language Models (LLMs) typically requires embedding entire tables into prompts. This approach faces challenges such as context window limitations, high computational costs and data leakage risks, particularly for large tables. To address these issues, we propose DataFrame QA, a task and framework that generates Pandas queries for information retrieval and analysis on tables. By using only table column names and data types, this approach ensures data privacy, reduces token usage and enhances efficiency, providing a foundation for secure and scalable LLM-powered tabular data analysis.



Machine Learning, Deep Learning and Reinforcement Learning

Our research spans a wide range of topics in machine learning and deep learning, including machine learning theory, reinforcement learning, graphical neural networks, adversarial learning, interpretable machine learning, federated learning and natural language processing. Our primary goal is to advance the state-of-the-art in machine learning theory and applications, addressing real-world challenges across various industrial and scientific domains. Our application fields include FinTech, Blockchain, transportation, urban computing, bioinformatics, neuroscience, social network analysis and cyber-physical system design. Many machine learning faculty members are also core faculty of the Center for AI Research (CAIR). Our research has received broad support by different agencies, including NSF, DOE, AFOSR, DOT, NIH, DOD and industry partners. The results are consistently published in high-impact journals and top-tier machine learning and artificial intelligence conferences, such as Nature Machine Intelligence, Nature Communications, IEEE Transactions on Neural Networks and Learning Systems (TNNLS), NeurIPS, ICML, KDD, AAAI, IJCAI and ICDM.



Jing Li
Associate Professor
jingli@njit.edu

Research Area: Real-time systems, parallel computing, cyber-physical systems and reinforcement learning for system design and optimization

Reinforcement Learning-Based System Design

The design space of modern complex systems is increasingly large. Finding good designs often involves solving mixed-integer optimization problems that are highly intractable. Our research develops reinforcement learning-based frameworks with graph neural networks and active learning techniques to intelligently and efficiently find good designs from the huge design space. We have applied our frameworks to various system design tasks, including resource allocation in cyber-physical systems and circuit design.

Akshay Rangamani
Assistant Professor
akshay.rangamani@njit.edu

Research Area: Deep learning, signal processing, computational neuroscience, neural collapse

Sparse and Low Rank Structures in Deep Networks

Deep learning has demonstrated great success across tasks in diverse domains that involve understanding image, audio and text. However, a mathematical description of how and why deep learning works and how to reliably generate models that generalize well is still incomplete. We aim to provide a description of how deep networks learn sparse and low rank features in their layers and how these structures can be used to provide generalization guarantees for deep networks. Neural Collapse is one such low rank structure among others that can provide this description. This project aims to characterize the conditions under which Neural Collapse emerges and how it can provide generalization guarantees.

Fine grained control of Multimodal Models through Neural Collapse

Neural Collapse (NC) is an emergent phenomenon of training deep networks that describes low rank structures that emerge in deep network layers. We will use the tools of NC analysis to improve the supervised fine-tuning step of training multimodal models. NC analysis can pinpoint the layers to be adapted, hence saving memory and time. It can also accelerate Supervised Finetuning (SFT) by guiding training towards the desired classifier geometry. We will also explore how Neural Collapse geometry can be used in two contexts: 1) continual learning of new concepts and 2) unlearning of concepts that need to be deleted. Identifying simple structures in these models through NC geometry can help accelerate these tasks.

Usman Roshan
Associate Professor
usman.w.roshan@njit.edu

Research Area: Machine learning, medical informatics

Adversarial Robust Machine Learning With 0-1 Loss

Machine learning models today are highly accurate but not very robust. They can be fooled to misclassify data with minor perturbations known as adversarial attacks. Adversaries targeting a given convex model are known to affect other convex models. We find this transferability phenomenon to be less effective between 0-1 loss and convex losses such as hinge and logistic, both of which are approximations to 0-1 loss and known to be affected by outliers. Consequentially, it is harder to attack 0-1 loss models with convex substitute model black-box attacks and when the black-box attacker is 0-1 loss, the attack is highly ineffective on all models. Based upon these observations, we are researching novel algorithms and design implementations for scalable and faster 0-1 loss models.

Hai Phan
Associate Professor
hai.phan@njit.edu

Research Area: Social network analysis, machine learning, spatio-temporal data mining

Ontology-Based Interpretable Deep Learning

Machine learning models are trained with large amounts of data and achieve a certain level of competency in interpreting and classifying new input data. However, even if they work well, it can be difficult to explain why. Lingering doubt persists that, in some situations, the classification output of the model might be wrong. In applications such as self-driving cars, this could have spectacularly negative consequences. We tie predictions of the model to a set of keywords taken from a predefined vocabulary of relevant terms. The number of words hard-coded into the model that influence the outcome produced by a machine learning for a new input is reduced and those words are taken from a limited and relevant ontology. This makes the output of the model easier to interpret as it becomes independent from terms that are irrelevant to the application domain.

Guiling "Grace" Wang
Distinguished Professor
gwang@njit.edu

Research Area: Applied Deep learning, AI in Finance, AI in Transportation, LLM

Deep Reinforcement Learning for Intersection Control

Reinforcement Learning-based traffic signal controllers can adaptively adjust signals based on real-time demand. To learn a good policy, we have proposed combining a dueling network, target network and double Q-learning network with prioritized experience replay technology in one signalized intersection. It has proven to be a successful attempt to stabilize the learning process and mitigate over-estimations of the learning agent. To further guarantee the safety of vehicles, we have incorporated domain safety standards in the above RL-based traffic signal controller. Nonetheless, the small proportion of the collision data makes this problem extremely challenging, not to mention that the learning agent might not obey the safety rules in practice. We thus, instead of letting the RL model learn by itself, incorporate domain safety standards into different parts of the RL model (i.e., action, state, loss function and reward function) as a safety shield. This safety-enhanced approach can guide and correct the RL agent toward learning a much safer action and dramatically lower the collision rate. Meanwhile, we not only focus on one intersection but also consider multiple intersections as a network to make a smoother driving experience. By enabling cooperation among intersections, each RL agent learns to communicate with others to spread out the traffic pressure quickly. Intensive experiments show the effectiveness of our systems.

Mengjia Xu
Assistant Professor
mx6@njit.edu

Research Area: Machine learning, LLMs for dynamical systems (e.g., climate modeling), causal inference, AI for Neuroscience

Fully Hyperbolic Graph Neural Networks for Brain Aging Trajectory Detection

Characterizing age-related alterations in brain networks is crucial for understanding aging trajectories and identifying deviations indicative of neurodegenerative disorders, such as Alzheimer's disease. In this study, we developed a Fully Hyperbolic Neural Network (FHNN) to embed functional brain connectivity graphs derived from magnetoencephalography (MEG) data into low dimensions on a Lorentz model of hyperbolic space. Using this model, we computed hyperbolic embeddings of the MEG brain networks of 587 individuals from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) dataset. Notably, we leveraged a unique metric—the radius of the node embeddings—which effectively captures the hierarchical organization of the brain, to characterize subtle hierarchical organizational changes in various brain subnetworks attributed to the aging process. Our findings revealed that a considerable number of subnetworks exhibited a reduction in hierarchy during aging, with some showing gradual changes and others undergoing rapid transformations in the elderly. Moreover, we demonstrated that hyperbolic features outperform traditional graph-theoretic measures in capturing age-related information in brain networks. Overall, our study represents the first evaluation of hyperbolic embeddings in MEG brain networks for studying aging trajectories, shedding light on critical regions undergoing significant age-related alterations in the large cohort of the Cam-CAN dataset.

Towards Efficient Edge-Aware Dynamic Graph Embedding with Mamba

Dynamic graph embedding is crucial for modeling time-evolving networks across various domains. While transformer-based models capture long-range dependencies in temporal graphs, they struggle with scalability due to quadratic complexity. Our study compares transformers with the Mamba architecture, a state space model with linear complexity. We developed three hybrid models: TransformerG2G with graph convolutional networks, MambaG2G and MambaG2G enhanced with graph isomorphism network edge convolutions. Experiments show that Mamba-based models offer comparable or superior performance to transformers in link prediction tasks while being more computationally efficient, especially on longer sequences. MambaG2G consistently outperforms transformers on variable datasets like UCI, Slashdot and Bitcoin and remains competitive on stable graphs like SBM. Additionally, Mamba-based models provide interpretable insights by analyzing attention weight and state matrices, advancing efficient temporal graph representation learning for applications like climate modeling, finance and biological systems.

Lingxiao Wang
Assistant Professor
lw324@njit.edu

Research Area: Privacy and Security in Machine Learning, Collaborative Learning (Distributed and Federated Learning), Optimization for Machine Learning, Deep Generative Models (e.g., Diffusion Models)

Towards the Next Generation of Federated Learning System for Foundation Models

Foundation models (FMs), such as ChatGPT and Diffusion Models, have driven significant breakthroughs across various fields, particularly in text and image generation. The success of these FMs depends on extensive training costs and large amounts of data. However, modern data generation — from individuals' personal devices, in smart homes and cities, within hospitals or financial institutions — fundamentally changes machine learning pipelines from classical scenarios, where we view data as a sample from a single large underlying population. These new data modes result in heterogeneous siloed data residing in the devices or organizations that generated it.

Such a shift presents new challenges in the development and deployment of FMs, such as improving the accuracy and efficiency of training across siloed data, mitigating risk and protecting data privacy and ownership and incorporating social and economic principles that incentivize data sharing and provide trustworthy cooperative learning schemes. Our research aims to develop the next generation of federated learning systems that enables coordinated, efficient, secure and trustworthy training of FMs across multiple parties and diverse data sources.

Jason Wang
Professor
wangj@njit.edu

Research Area: Data mining, machine learning, deep learning, explainable AI, generative AI, trustworthy AI, data science

Cyberinfrastructure-Enabled Interpretable Machine Learning

Cyberinfrastructure (CI) enabled machine learning refers to new computing paradigms such as machine-learning-as-a-service (MLaaS), operational near real-time forecasting systems and predictive intelligence with Binder-enabled Zenodo-archived open-source machine learning tools, among others. These computing paradigms take advantage of advances in CI technologies, incorporating machine learning techniques into new CI platforms. In this project we focus on interpretable machine learning where we attempt to explain how machine learning works, why machine learning is powerful, what features are effective for machine learning and which part of a testing object is crucial for a machine learning model to make its prediction. Methods, techniques and algorithms developed from this project will contribute to advancements of CI-enabled predictive analytics and explainable artificial intelligence in general.

Lijing Wang
Assistant Professor
lw29@njit.edu

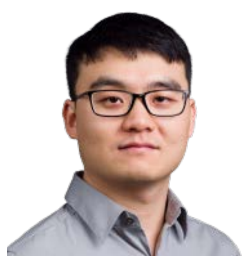
Research Area: Machine Learning, deep learning, epidemic modeling/forecasting, clinical NLP, ML in Social Science

Exploring Spatial, Temporal and Semantic Patterns for Graph Neural Networks

The project focuses on the exploration and integration of spatial, temporal and semantic patterns within the framework of Graph Neural Networks (GNNs). As GNNs have become a powerful tool for processing and learning from data structured as graphs, there is growing interest in extending their capabilities to better handle complex data that exhibits not only topological relationships but also dynamic temporal behaviors and rich semantic information. This project aims to advance the state-of-the-art in GNNs by investigating methods to effectively capture and utilize these three dimensions—spatial, temporal and semantic—to enhance performance across various applications, such as social network analysis, traffic prediction and recommendation systems.

Scalable and Robust AI Models for Information Extraction

Extracting information from EHRs has been an active area of research in recent years due to the advances in natural language processing (NLP) techniques. Large language models (LLMs), such as Bidirectional Encodings Representations from Transformers (BERT), have achieved state-of-the-art performance for a variety of NLP tasks through either pretraining a domain-specific LLM from scratch or fine-tuning a general domain pretrained LLM on a domain-specific dataset. Our research will investigate advanced scalable and robust LLM architectures for improved generalization, adaptability and efficiency for information extraction from EHRs and general text information.



Shuai Zhang
Assistant Professor
sz457@njit.edu

Research Area: Parameter-efficient transfer learning, LLMs, Weight analysis, LoRA, Task vector, Machine unlearning, Multi-task learning

Parameter-efficient Transfer Learning

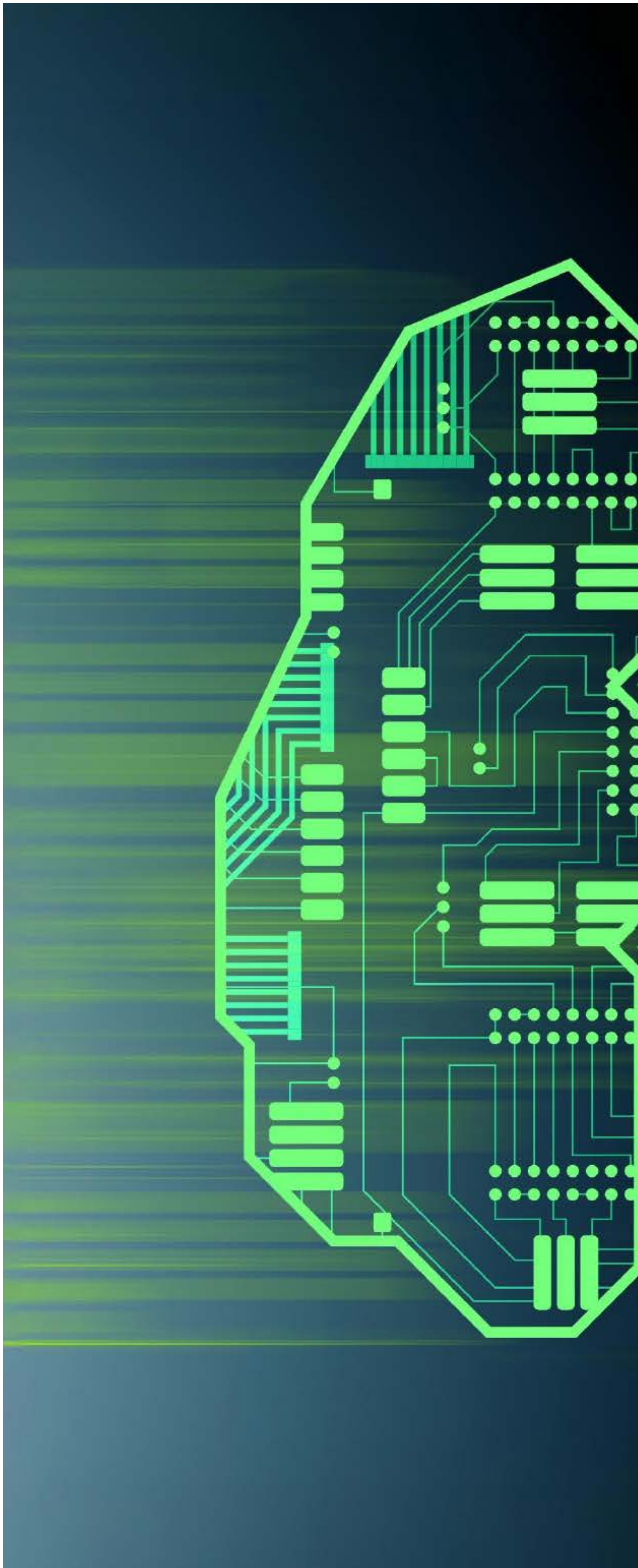
Large pre-trained models have become foundational modules in modern machine learning systems. In the pre-training and fine-tuning paradigm, traditional full-parameter fine-tuning demonstrates superior performance but results in significant computational inefficiency. Our goal is to systematically analyze the attributes of the weights in pre-trained models with theoretical guarantees and design a parameter-efficient transfer learning approach to adapt models to specific requirements. To meet the diverse dimensional requirements of pre-trained models, we also aim to design a modular architecture that allows different components to be integrated orthogonally, enabling rapid adaptation without the need for extensive fine-tuning.

Zhi Wei
Distinguished Professor
zhiwei@njit.edu

Research Area: Machine learning, statistical modeling, bioinformatics

Explainable AI for Unsupervised Learning

In recent years, explanation techniques have emerged and received a lot of attention. In supervised learning settings, it emphasizes the ability to correctly interpret a prediction model's output. Most existing explanation works have been focused on supervised learning, such as LIME and SHAP. Yet very little work has been done in an unsupervised learning setting. Our goal is to explain why and how a sample is assigned to a specific cluster in unsupervised learning. Specifically, we would like to know which features are portrayed as contributing to a cluster, or evidence against it. With this information, a practitioner can make an informed decision about whether to trust the model's cluster assignment. There is also a so-called "double use of data" problem when trying to find discriminative features that distinguish the resultant clusters. We will apply the new methods to analyze finance and accounting data.



Mobile Computing

Mobile computing and sensing have become ubiquitous in the world, due to the widespread availability of smart phones, smart watches and IoT devices. The availability of computing and sensing anytime, anywhere presents opportunities for novel applications that can improve our daily life, in domains such as mobile health, transportation, communication and safety. We focus on applying machine learning techniques to solve mobile computing and sensing challenges, as well as designing collaborative intelligence solutions for human-robot teams. The results of this research have been published in top venues such as AAAI, ECAI, IEEE/RSJ IROS, IEEE MASS and IEEE Transactions on Mobile Computing.

Cristian Borcea
Professor
borcea@njit.edu

Research Area: Mobile systems, federated learning

Privacy-Preserving Federated Multi-Task Learning

Multi-task learning (MTL) enables simultaneous learning of related tasks, enhancing the generalization performance of each task and facilitating faster training and inference on resource-constrained devices. Federated Learning (FL) can further enhance performance by enabling collaboration among devices, effectively leveraging distributed data to improve model performance, while ensuring that the raw data remains on the respective devices. However, conventional FL is inadequate for handling MTL models trained on different sets of tasks. We propose FedMTL, a privacy-preserving FL aggregation technique that handles task heterogeneity across users. FedMTL generates personalized MTL models based on task similarities, which are determined by analyzing the parameters for the task-specific layers of the trained models. To prevent privacy leakage through these model parameters and to protect the privacy of the task types, FedMTL employs low-overhead secure multi-party computation aggregation.

Kasthuri Jayarajah
Assistant Professor
kasthuri.jayarajah@njit.edu

Research Area: Sensing, Robotics, Context Awarenesse, Mobile Systems, Edge AI

Human-Robot Cooperative Systems

Effective communication between humans and robotic teammates is imperative for successful cooperation. However, direct communication (e.g., verbally) can be distracting to humans under high cognitive workload scenarios. Thus, we explore the feasibility of using multimodal wearable sensing to facilitate implicit coordination where robots are able to synchronize with their human teammates without explicit intervention. We have developed systems that extract human intent and task performance using physiological sensing, visual attention and gestures for cooperation in scenarios such as search and rescue, teleoperation and cognitive rehabilitation, addressing key system challenges in dealing with varying sensing fidelity, hardware resource constraints and stringent latency bounds.

Collaborative Edge Intelligence

While collaborative intelligence over a network of sensors can lead to more accurate inferencing as opposed to isolated, individual inferences, the cost of collaboration (e.g., increased bandwidth requirement) can be highly prohibitive. The problem is especially exacerbated for richer data representations such as images and point clouds. We developed various architectures for collaboration for multi-view camera systems and heterogeneous multi-robot systems investigating the performance trade-offs of early and late fusion methods and more recently, Com AI, where we adapt deep learning pipelines on-the-fly based on lightweight digest sharing between collaborating nodes. ComAI incurs zero re-training of the deep networks, adds minimal overhead to the processing pipeline as well as the bandwidth required for sharing digests over the network. We further explore self-configuration and steerability for collaborative systems using Reinforcement Learning (RL) where cameras learn to identify regions that benefit the most from collaboration, quantify the goodness of collaborators for selective collaboration and over time, "steer" themselves towards optimal configurations.

Privacy

Our research focuses on enhancing privacy and security in sensitive data applications, including healthcare and digital advertising. Through advanced protection systems and regulatory compliance studies, we aim to safeguard personal information and promote secure, privacy-compliant digital ecosystems. Our work has been published in top venues, including the ACM Conference on Computer and Communications Security.

Hai Phan
Associate Professor
hai.phan@njit.edu

Research Area: Social network analysis, machine learning, spatio-temporal data mining

Preserving Sensitive Health Care Data

Networks that share electronic health records containing a patient's personal and clinical data promise improved continuity of care and better health outcomes. However, these networks put highly sensitive patient information at risk and expose healthcare providers to legal jeopardy. We created DeepPrivate, a system that uses machine learning techniques to protect personal health information against cyber attacks.

Cristian Borcea
Professor
borcea@njit.edu

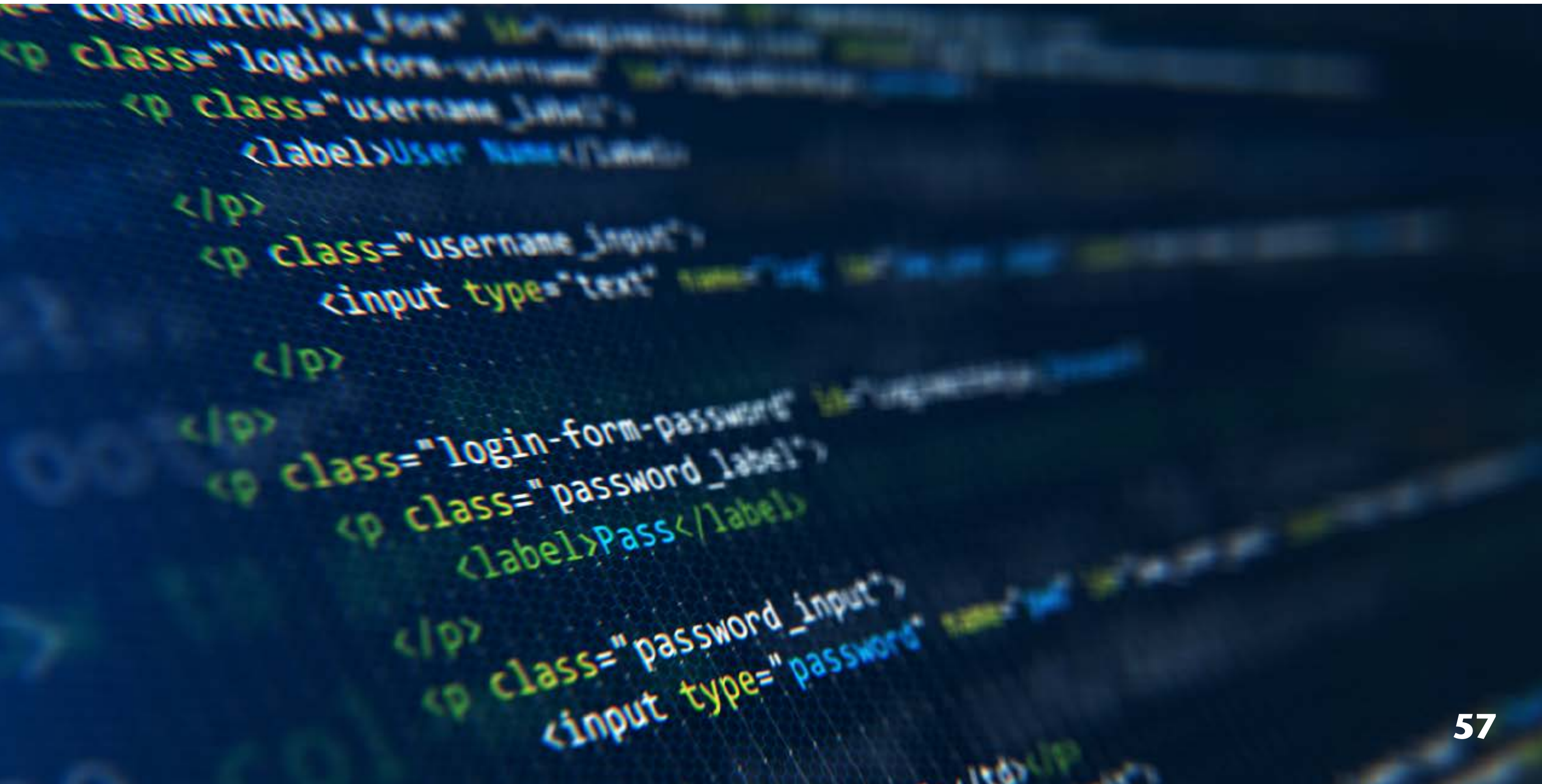
Research Area: Privacy, online advertising

Online Publishers Compliance with Privacy Regulations

Privacy regulations are enacted by governments around the world to protect their citizens. The most well-known such regulation is the General Data Protection Regulation (GDPR), for protecting European Union (EU) citizens from unnecessary and unauthorized personal data collection. We studied GDPR compliance under the Transparency and Consent Framework (TCF), which is widely used across the Internet and provides digital advertising market participants a standard for sharing users' privacy consent choices. We found that most websites properly record the user's consent choice, but over 72% of the websites that were TCF compliant claimed legitimate interest as a rationale for overriding the consent choice. While legitimate interest is legal under GDPR, using it at large scale contradicts the spirit of the law. Additionally, analysis of cookies set to the browsers indicates that TCF may not fully protect user privacy even when websites are compliant.

Software Engineering and Programming Languages

Software pervades all aspects of modern life; unfortunately, so do software failures. In other engineering disciplines product correctness and reliability are the norm, but in software engineering failures are a fact of life. The price tag of this inadequacy is massive: A study by the Consortium for Information and Software Quality estimated that, as of 2020, the cost of poor quality software in the USA exceeds $2.3 trillion. Like all complex, multi-dimensional problems, the challenge of software quality cannot be addressed by any single approach. Our software engineering research includes a rich, diverse portfolio of approaches (analytical, empirical), paradigms (static, dynamic), quality attributes (correctness, reliability, security), lifecycle phases (testing, verification, repair) and products (methods, tools, artifacts).



Ali Mili
Professor
mili@njit.edu

Research Area: Software engineering, software testing, program correctness, program repair

Functional Extraction

Despite several decades of research and technology transfer, the development of reliable and secure software remains an unfulfilled goal: software is routinely fielded with known failures but unknown faults. Many of the problems of software engineering stem from the fact that programmers develop and reuse software products whose function they do not fully comprehend. The purpose of the functional extraction project is to develop and evolve an automated tool that computes the function of a program by static analysis of its source code in a C-like language. We achieve this goal by mapping a program onto an equation between its inputs and outputs, which we can then solve or reason about to analyze correctness, safety, security, reliability, etc. The main innovation of this project is the ability to compute the function of loops, which we achieve by means of invariant relations.

Theoretical Foundations for Program Repair

To repair a program does not mean to make it correct; it only means to make it more-correct, in some sense, than it was. In this project we introduce a definition of relative correctness, a reflexive transitive partial ordering between candidate programs that culminates in absolute correctness with respect to a reference specification and we show how this concept can be used in program repair to: define what is a fault; define what is a unitary/atomic fault; define what it means to repair a fault; measure the degree of faultiness of a program; and perform program repair in a stepwise manner, by removing one unitary fault at a time until the program is absolutely correct. We show empirically that by augmenting existing program repair tools with our capabilities, we enhance their performance and their scope.

Mutation Testing

Mutation testing is the art of generating mutants of a base program and is used primarily to assess the quality of test suites: a test suite is all the better that it can distinguish the base program from its non-equivalent mutants. We estimate the ratio of equivalent mutants of a base program by analyzing the amount of redundancy in the program. Also, we analyze the effectiveness of a mutant set by its ability to vet test suites that expose the failures of the base program; and we model the minimization of a test suite as an optimization problem, where the objective function is the cardinality of the set and the constraint under which the optimization is sought is the preservation of the mutant set effectiveness.

Failure Based Test Suite Effectiveness

The most important step we take in software testing is the selection of test data and the most important factor in this step is the way we assess and compare the effectiveness of test suites. Most existing metrics to assess the effectiveness of test suites equate effectiveness with the ability to detect faults; but we argue that there is another way to assess the effectiveness of a test suite, by its ability to expose failures. In this project we explore the implications of this dichotomy.

Martin Kellogg
Assistant Professor
martin.kellogg@njit.edu

Research Area: Verification, automated reasoning, software engineering, programming languages

Inferring Type Qualifiers

Pluggable type systems enable programmers to enhance the ability of a programming language's type system to automatically prevent bugs by attaching a type qualifier to each type in the program: each pluggable type system prevents one kind of bug. These techniques have been shown to be effective, but can be difficult to apply to legacy systems because of the need to write type qualifiers. We have developed a novel type inference system that is general, meaning that it applies to any pluggable type system without needing to modify the pluggable type system itself, making the techniques more accessible to programmers.

Correlating Understandability and Verifiability

Intuitively, code that is simple for a human to understand ought to be simple to automatically verify as free of bugs. We have developed an experimental methodology to test for this correlation in real code and applied it to a corpus of programs used in prior studies of how humans understand code and found evidence for this intuition. Our findings have implications for how we build and use verification tools: when automated tools fail to verify code, it does make sense to refactor that code to make it simpler, because code that is simpler for the verifier to understand is more likely to be simple for other programmers to understand as well.

Iulian Neamtiu
Professor
iulian.neamtiu@njit.edu

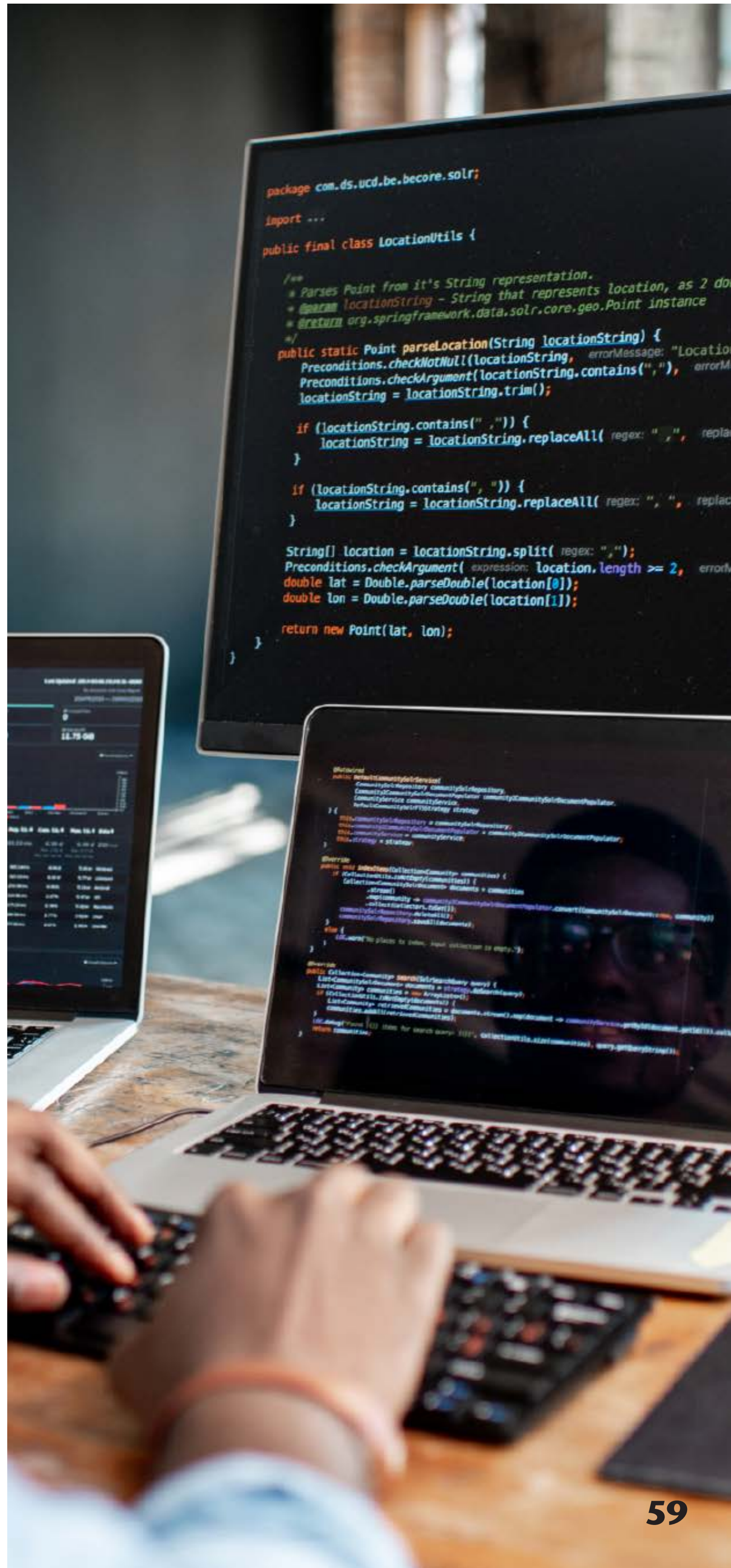
Research Area: Programming languages, software engineering and their applications to reliable AI, smartphones, security

Improving Smartphone Reliability and Security

We have developed a wide range of approaches to improve smartphone reliability, including static and dynamic program analyses, record-and-replay systems, application self-healing and automatic test generators. These approaches were released as open-source code and have exposed reliability issues in many popular applications including medical mobile apps used in acute care settings.

Verifying Unsupervised Learning Implementations

Unsupervised Learning (UL) is widely used, by itself or as part of AI pipelines, in scenarios ranging from everyday applications to high-stake domains. However, even specifying UL correctness is a challenge, which complicates verifying or validating UL implementations. Our past work has exposed serious issues (e.g., nondeterminism, inconsistency) in widely popular UL toolkits and on critical, e.g., medical datasets. Our current work is focused on automated verification and fault localization for UL implementations using static and dynamic analysis.




NJIT
1881
NJIT
Computing
NJIT
Honors College
NJIT
Management
NJIT
Science
& Liberal Arts
NJIT
Architecture
& Design
Engineering

NJIT | YING WU COLLEGE OF COMPUTING

Ying Wu College of Computing
New Jersey Institute of Technology
University Heights
Newark, NJ 07102

973-596-2822 | computing.njit.edu