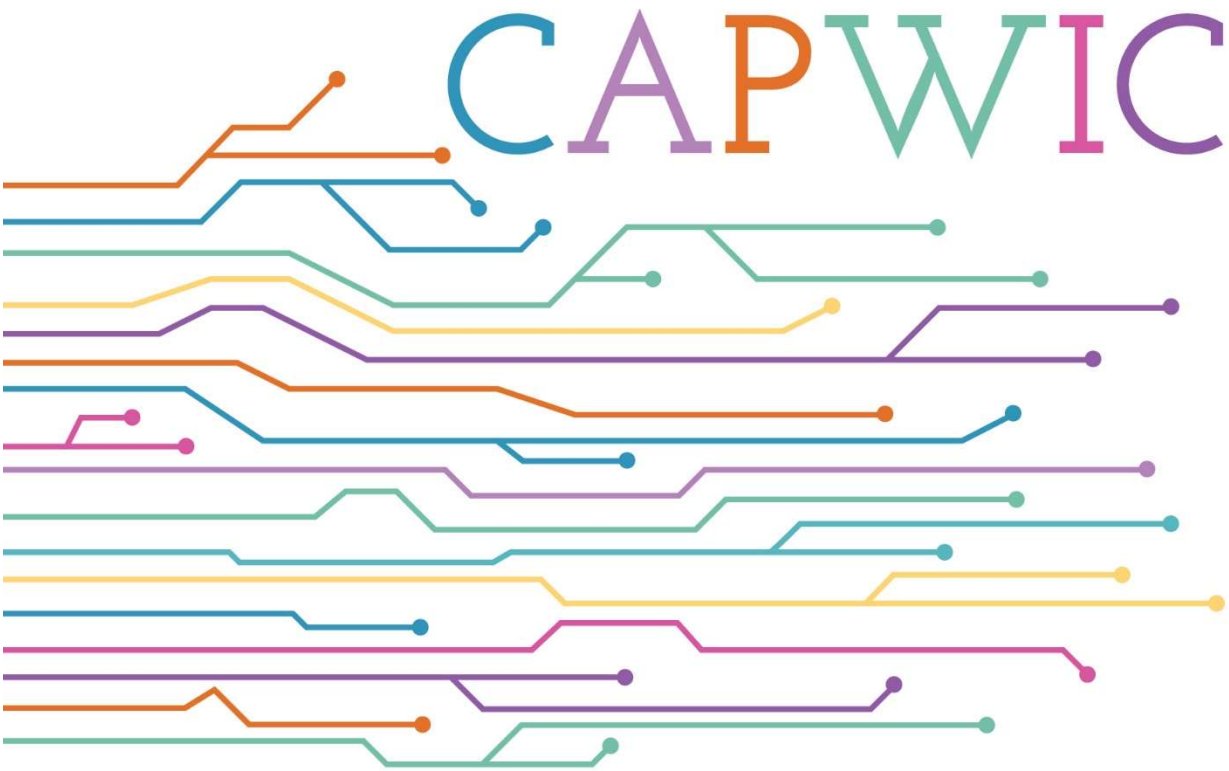
CAPWIC 2021

The ACM Capital Region Celebration of Women in Computing

Abstracts listed in alphabetical order by first author.

Contents



Flash Talks

Meenakshi Balan (Georgetown University).

Characterizing the role of newspapers in conversations around the 2020 election candidates.

Abstract. This project attempts to characterize the role of traditional media in the 2020 election by investigating the relationships between media bias, newspaper endorsements, and circulation size. Newspapers have historically played a formative role in shaping the conversations around presidential elections; the shift of conversations towards social media and questions about media bias suggest that the function of newspapers may be evolving. Using Python we applied a dictionary-based sentiment method to 72,762 articles pertaining to either candidate from over 700 newspapers. We then segmented the newspapers into three subgroups: (1) we isolated the "Big 10" newspapers, which account for the largest average weekday circulation, (2) we separated liberal, neutral, and conservative newspapers using ratings from Media Bias/Fact Check, and (3) we identified newspaper endorsements across the 2016 and 2020 election cycles. After analyzing the volume of articles and sentiment scores across these three subgroups, we observed a few trends: that the "Big 10" and liberal categories contributed the highest volume of articles in our sample, that "Trump''-titled articles accounted for a higher volume than "Biden"-titled ones, and that "Biden"-titled articles were correlated with more positive sentiment rankings than "Trump"-titled ones. The greater representation of liberal-leaning news over conservative-leaning news suggests that the high-circulation newspaper landscape may tend more towards liberal or moderate points-of-view, perhaps overshadowing other conversations occurring at the local level. In future research, we hope to apply topic modeling to better understand key topics and the role that newspapers play in their surrounding conversations.

Cassie Bedard (James Madison University), Garrett Hutson (James Madison University), Evelyn Munstermann (James Madison University), Katelyn Anderson (James Madison University) and Shraddha Joshi (James Madison University).

Connected Relationships using Internet of Things.

Abstract. The goal of our 2 year long capstone project is to aid the older generation, ages 60-80, in enhancing connections with their grandchildren. We aim to do this by using the internet of things (IoT) technology to create an easy-to-use product that helps mitigate the disconnection that comes with being separated from loved ones.

Today's world has brought digital connection and physical distance to a new light. Due to obstacles such as lack of time or long distance, more and more people are using the virtual environment to stay connected to friends and family in everyday life, therefore our team's goal is to use IoT to augment relationships for people who can not be physically present.

In the fall 2020 semester we completed phase 1 focused on project exploration. We defined our stakeholder and focus for the project and conducted stakeholder interviews. Preliminary research led to the selection of grandparents and grandchildren due to reasons such as loneliness, health risks for grandparents, and the desire to bridge the technology gap between the two generations. Currently in the spring 2021 semester, we are using the most common disconnects, such as the inability to hug or spoil grandchildren, to develop potential concepts with the hope to decide on a concept that we believe to be most beneficial to our stakeholders. After a concept is chosen, our team can progress in

the design process to testing and fabrication. Fabrication, testing and improving design will be the goal of the remaining two semesters.

Pamela Bogdan (Ocean County College) and Sylvia Riviello (Ocean County College).

CSIT XXX – Applied Programming for All Majors: A Contextualized Course Based on Inter-Disciplinary Curriculum Development.

Abstract. This flash talk will outline our plans to create an entry level programming course that will appeal to all majors. It will be done using contextual learning techniques that will combine the topics of programming with application assignments that are tightly linked to each student's major. The course will be an inter-disciplinary curriculum development effort.

Darshea'L Cook (Virginia State University) and Joseph Shelton (Virginia State University).

Hands on Cyber Security Education using MITRE's CALDERA.

Abstract. One of the biggest issues in education is memorization. Many students have problems with flat-out memorizing and recalling information learned strictly from lectures. CALDERA is Automating Adversary Emulation where it teaches the user ATT&CK tactics and techniques to use on their networks. Adversary Emulation is the idea of creating a realistic adversary on a network where the user can control and monitor that adversary. This educational tool helps to identify the steps one needs to take to improve their systems from adversaries. We will create an interactive game that breaks down and provides step-by-step instructions for the player. The idea of this game uses similar tactics as Battleship or Chess with opponents battling against one another. The effects are improvement of the student's reaction time, decision-making skills, and overall understanding of cybersecurity. Using gaming in education to help combat cybersecurity has endless benefits and possibilities. For instance, many students have different learning styles providing each student with several of those styles will help education and the learning process. Though this study is still being tested, the goal for determining where each student stands will present itself in a pretest, which will determine the student's level of knowledge and skill-based. The measurement of learning throughout the game will come from sub quizzes in each section. By the end of the game, the student will then complete a post-test to determine the effectiveness of the game. CALDERA will be used to train and educate through an interactive strategic thinking pattern.

Samantha Dies (Georgetown University), Autumn Toney (Georgetown University) and Lisa Singh (Georgetown University).

Social Media as a Driver for Social Change: Analyzing the #MeToo and #BlackLivesMatter Movements on Twitter.

Abstract. With the significant increase of social movements activism online, it is important to understand the role online activism has in legal, social, and cultural change. Individuals from around the world are able to reach a global audience by sharing their narratives, demanding reform, and promoting protest events on social media platforms. Two movements that have gained considerable visibility on social media are the #MeToo and #BlackLivesMatter movements. This study presents a frequency analysis of the #MeToo and #BlackLivesMatter social movements on Twitter during critical 3-month periods for each movement respectively.

We analyze #MeToo between July and October, 2019 and #BlackLivesMatter between May and August, 2020. This analysis is the basis for determining the relationship between online activism, offline action, and ultimately, policy change. Using each movement's respective hashtag (#MeToo and #BlackLivesMatter), we compute the most frequent words and hashtags present in 2,808,678 #MeToo and 18,813,718 #BlackLivesMatter tweets. Referencing topologies generated for each movement, we categorize the most frequent words and hashtags to characterize the social movements and compare their online conversations. This characterization allows us to interpret what the majority of a social movement's tweets are responding to. Specifically, the topologies characterize tweets with respect to the general social movement and the drivers of social change, cultural change, or legal change. In our study, we find that more tweets focus on systemic structural change when discussing #BlackLivesMatter than #MeToo, and both movements have larger amounts of conversation related to awareness and activism.

Kristina Kramarczuk (University of Maryland College Park).

Secondary Science Teachers' Intentionality of Using Computational Thinking to Create Learning Opportunities for BIPOC Students.

Abstract. Schools with a majority of students eligible for free/reduced lunch are the least likely to have access to computer science (CS) courses. Integrating computational thinking (CT) into the science classroom is one way to increase access to CS. However, very little research explores how the racial and gender identities of practicing science teachers at Title I schools influence their perceptions and implementation of CT. In this study, I use Kimberlé Crenshaw's (1990) Intersectional Feminism lens to explore how two secondary science teachers of color (a Black woman and a Latina) define CT in the contexts of their Title I classrooms and how their racial and gender backgrounds influence their definitions. Semi-structured qualitative interviews revealed that the educators' racialized and gendered experiences influence how they define CT and that CT is not as widespread as reformers might hope. Both teachers agreed that CT is important and that their students should have access to CT curriculum, but they 1) were unfamiliar with the term before the interview and 2) believe that the marginalizing and exclusive practices of the mainly white and masculine STEM-CS community make it difficult for their students to identify with STEM-CS. However, both educators also use their personal experiences with discrimination in STEM-CS as motivation to hold high expectations for their students and promote critical thinking in their classrooms. Preliminary results from this study suggest that the white and masculine norms of the STEM-CS community must be confronted before CT can effectively be implemented and reach all students.

Tania Lorido-Botran (Independent Researcher) and Muhammad Khurram Bhatti (Information Technology University).

RLSched: Adaptive Container Scheduling in Cloud Data Centers using Deep Reinforcement Learning.

Abstract. Cloud data centers rely on virtualization to run a diverse set of applications. Container technology allows for a more lightweight execution, in comparison with popular Virtual Machines. Efficient scheduling of containers is still challenging due to varying request arrival patterns, application-specific resource consumption and resource heterogeneity in physical servers. Besides, containers are also more prone to resource contention and performance interference. Cloud providers need to overcome these challenges with a goal in mind: maximize resource utilization to satisfy as many

requests as possible. First, we introduce modeling of the container scheduling problem as a Markov Decision process, with formulation for the action space, state space and reward function. We propose RLSched, a deep reinforcement learning-based (DRL) scheduler that is self-adaptive and automatically captures the resource usage dynamics in the data center. The scheduler is based on a decentralized actor-critic multi-agent architecture that enables for parallel execution and faster convergence. RLSched relies on an enhanced neural network network model with action shaping, which filters invalid actions and prevents the agent to fall into a sub-optimal policy. The proposed scheduler is compared against other state-of-the-art DRL methods on a simulated data center environment based on real traces from Microsoft Azure. Preliminary results show faster convergence and higher number of containers placed per session.

Toriyonnah Lundy (Virginia State University), Snaie Minott (Virginia State University) and Joseph Shelton (Virginia State University).

Combined Human Factors and A.I Approach for Detecting Manipulated video.

Abstract. Deepfakes can be characterized as manufactured audio or visual media created utilizing profound learning, a subfield of AI (ML), that seem, by all accounts, to be real and are regularly made with the expectation of beguiling crowds. Artificially created media broadly varies in specialized complexity and application, going from bad quality "modest fakes" to all the more high caliber "deep fakes," and can challenge and impact view of the real world. The advancement of engineered general media content isn't novel: Hollywood movie producers have utilized PC created symbolism (CGI) since the 1970s to briefly for the term of the film suspend incredulity among crowds. Advances in ML have made complex manufactured media less expensive and simpler to deliver (especially because of a multiplication of free and open source programming for creating deep fakes. Indeed, even mechanically unsophisticated entertainers are currently ready to make and circulate deep fakes.

Deep fakes manipulate videos or images by making people appear to be saying or doing things that they really did not. Our team has compiled a list of ways to detect if a video is a deep fake. There are certain things to watch for to be able to tell if a video is a deep fake, such as blurring or misalignment, unnatural eye movement, inconsistency in the audio or video, and unnatural body posture and body movement.

Zahra Rizvi (York High School).

The Role of Data Science in Combating COVID-19 Pandemic.

Abstract. The COVID-19 pandemic, also known as the coronavirus pandemic, is a global pandemic of coronavirus disease 2019 (COVID-19). The World Health Organization (WHO) declared the outbreak a Public Health Emergency of International Concern and a pandemic in early 2020. As of late February 2021, more than 112 million cases of COVID-19 have been reported worldwide, resulting in more than 2.5 million deaths; more than 88 million people have recovered. Data science is the study of data that involves developing methods of recording, storing, and analyzing data to effectively extract useful information. The goal of data science is to gain insights and knowledge from any type of data. Data science is playing a crucial role in the global response to the COVID-19 pandemic. It is used to monitor real-time disease outbreak, forecasting, and spotting real-time trends for governments, public and private institutes, and health organizations. The ongoing COVID-19 pandemic data, if recorded properly, can become a goldmine for combating future pandemics. It is poised to make key contributions in the battle against current and future pandemics, from monitoring and tracking

cases/deaths to getting valuable insights about how populations move during travel restrictions to vaccine design. In my flash talk, I will highlight the benefits of data science to strategists and policymakers, and how data science can guide them in coming to grips with the issues, challenges, opportunities, and downsides involved in using data science to combat the COVID-19 and/or future pandemics.

Donna Schaeffer (Marymount University) and Patrick Olson (National University).

AI Advancing Social Welfare During Pandemics.

Abstract. The Covid-19 pandemic has affected employment and the social net in countries around the world. Before Covid-19, the Organization for Economic Cooperation and Development (OECD) reported that the number of people world-wide who are either self-employed or employed on a part-time and/or temporary basis is rising. These people are less likely to be unionized and less likely to receive welfare services when out of work. This issue is exacerbated by the pandemic.

The OECD report calls for better social protection coverage for workers in non-standard jobs. Increasingly, social protection systems are integrating technology advances, such as Artificial Intelligence (AI). Since Covid-19, we see a proliferation of tracking applications that citizens are encouraged to use as well as more demand on the social net.

The integration of AI, and other technology advances, raise concerns on how to maintain personal and societal cybersecurity. These systems work with personal information from one's health records, employment information, and data from financial institutions. The systems even pull in information from family members.

In this flash talk, we discuss policies and legislation that can enhance cybersecurity and protect people's data, privacy, and well-being as technology advances, such as AI, become integrated into social welfare.

Weizhen Sheng (Percentage Project), Stephanie Tang (Percentage Project) and Lucia Huo (Percentage Project).

Exploring Data-Driven Advocacy.

Abstract. Do you feel intimidated studying Computer Science? Do you experience microaggressions? Do you have a professor/mentor whom you look up to? The Percentage Project, a 501(c)(3) nonprofit organization that focuses on data-driven advocacy, asks powerful questions like these in its annual climate survey for post-secondary school students.

Our 2019-20 survey found that 62% of students identifying as women (versus 27% men, 41% non-binary) agreed that they feel intimidated studying CS. We use this survey result alongside others to help inform the audience on the current climate of college computer science communities. We pair these statistics with an annual photography campaign on social media, presenting a "Humans of New York" approach to understanding the experience of computer science students today.

In our flash talk, we will present select statistical results from our annual surveys that are most insightful or thought-provoking, as well as share some highlights from our past annual social media campaigns. Ultimately, participants will walk away from the session with reflections about computer science studies in today's college communities.

Student Posters

Eleni Adam (Old Dominion University), Desh Ranjan (Old Dominion University) and Harold Riethman (Old Dominion University).

NPGREAT: A method for assembling the human subtelomeres with the use of ultralong Nanopore and Linked-Read datasets.

Abstract. The telomeres are the protective caps of the chromosomes. Possible rearrangements can result to the dysfunction of the telomeres and subsequently to a wide range of diseases. The telomeres and subtelomeres however, are regions that have not been studied thoroughly due to the absence of the necessary technology. Their repetitive and complex structure causes the difficulty for their assembly. We develop the NPGREAT (NanoPore Guided Regional Assembly Tool), a hybrid assembly method to overcome this problem and assemble the human subtelomeric regions accurately. The method uses two types of data, the ultralong Nanopore reads and the Linked-Reads. The ultralong Nanopore reads are used as scaffolds upon which the REXTAL contigs can be placed and corrected, replacing the low-quality Nanopore sequence with high-quality sequence for matching regions. We tested NPGREAT for the NA12878 human subtelomeres, obtaining for the first time high quality assembly with high percent identity when compared to the reference.

Maytha Alshammari (Old Dominion University) and Jing He (Old Dominion University).

Analysis of Secondary Structure Geometry Accuracy in Ab Initio Protein Structure Predictions.

Abstract. Protein structure prediction produces atomic models of three-dimensional structure of a protein from its amino acid sequence. Predicting the three-dimensional structure is an important task since the three-dimensional structure of a protein provides unique information about the function mechanism of the protein. Here we analyze models predicted using five ab initio protein structure prediction methods (AlphaFold2, Baker, Quark, Zhang-Server, and MULTICOM) in protein structure prediction at Critical Assessment of protein Structure Prediction (CASP 14). The analysis is focused on the secondary structure geometry, the length, and the relative position of the predicted secondary structure. We found that the current advanced methods are performing well in predicting most secondary structures (α-helices and β-strands) of a protein. However, in some cases, the methods were not able to accurately predict the relative geometry between secondary structures which might affect the overall folding relationship among secondary structures. We propose a potential that using information from Cry-EM images to enhance the ab initio protein structure prediction by providing accurate geometry statistics among the protein secondary structures.

Hanjie Chen (University of Virginia), Yangfeng Ji (University of Virginia) and Hanjie Chen (University of Virginia).

Improving the Explainability of Neural Sentiment Classifiers via Data Augmentation.

Abstract. Sentiment analysis has been widely used by businesses for social media opinion mining, especially in the financial services industry, where customers' feedbacks are critical for companies. Recent progress of neural network models has achieved remarkable performance on sentiment

classification, while the lack of classification interpretation may raise the trustworthy and many other issues in practice. In this work, we study the problem of improving the explainability of existing sentiment classifiers. We propose two data augmentation methods that create additional train-ing examples to help improve model explainability: one method with a predefined sentiment word list as external knowledge and the other with adversarial examples. We test the proposed methods on both CNN and RNN classifiers with three benchmark sentiment datasets. The model explainability is assessed by both human evaluators and a simple automatic evaluation measurement. Experiments show the proposed data augmentation methods significantly improve the explainability of both neural classifiers.

Ana Valentin (Marymount University) and Sharon Lee Reagan Diaz (Marymount University).

Development of a project proposal applying project management principles.

Abstract. Companies apply project management practices to reduce project costs, improve job performance, and deliver the product on time. In this poster, the researchers present a project proposal blueprint applying the principles of project management for the dissertation project process. The development of the two-page project proposal illustrates a roadmap of a computer science project proposal's that laying out the purpose, goals, objectives, stakeholders, deliverables, and milestones of the project. The project proposal includes identifying and defining the roles and responsibilities of stakeholders and resources needed to complete the project proposal. Achieving a project proposal demonstrates that you plan to accomplish the project within scope, schedule, and budget. The majority of doctoral graduate schools require a project proposal as part of the admission package to assure graduate candidates complete the dissertation projects in a timely manner.

Katie Zhang (Virginia Commonwealth University) and Carol Fung (Virginia Commonwealth University).

Privacy Concerns Pertaining to Smart Home Devices.

Abstract. Over the years, technology has improved significantly and integrated into people's daily lives. People have become dependent on technology, specifically smart home devices (SMDs), for day-to-day activities. However, privacy concerns are now a major issue because of SMDs' access to people's personal information. Various studies have been conducted on the privacy risks posed by SMDs, but few address the problem from a user-centered perspective. In our work, we seek to uncover user perceptions of SMD privacy and provide insight on ways researchers can raise awareness of affiliated privacy concerns. In an 18-question online survey, 1000 participants answered questions about their history with, perceptions of, and concerns related to SMDs. A series of six questions evaluated user ability to select essential and non-essential data.

Our data analysis found that people scored an average of 73.8% on the accuracy test that judged ability to differentiate between essential and non-essential data. Out of six questions, this percentage is equivalent to answering roughly 4.4 questions correctly. The vast majority of users expressed concern for privacy leaks from SMDs, and the top privacy concern is the leak of username/password. However, most people are still willing to sacrifice privacy for functionality. Our work helps researchers better understand public perception of SMDs and develop effective methods of educating the public on information privacy, as well as teach people how to protect their privacy. Further research can be conducted to determine how people would best respond to efforts to raise awareness of privacy concerns surrounding SMDs.

Student Research Shorts

Bipasha Banerjee (Virginia Tech), William A. Ingram (Virginia Tech) and Edward A. Fox (Virginia Tech).

Extracting Information from Electronic Thesis and Dissertations.

Abstract. Theses and dissertations are long documents that record graduate students' work. Today, many universities require students to submit an electronic thesis or dissertation (ETD) as a graduation condition. ETDs are rich sources of knowledge and often contain cutting-edge research findings. Virginia Tech, a leader in the ETD movement now has a collection of over 33 thousand documents. In 2019, Virginia Tech and Old Dominion University were awarded a 3-year National Digital Infrastructures research grant from the Institute of Museum and Library Services (LG-37-19-0078-19) to study the application of computational methods on large corpora of ETDs. Our work focuses on using computational techniques and algorithms for classification, segmentation, summarization, and extraction to facilitate the identification, discovery, and reuse of the knowledge contained in ETDs.

Existing digital libraries can be used to find such documents; however, an ETD is less useful without guidance as to which portion is most relevant to the information need. ETDs are organized into chapters and each chapter can be conceptualized as the equivalent of a research article. However, unlike a research article that typically has an abstract, it is uncommon for ETD chapters to have their own abstracts. Hence, one of the main goals of my research is to produce a condensed summary for each chapter of an ETD, along with other key information mined from the document. My preliminary research, i.e., investigating and implementing existing approaches to segment and then summarize the identified parts of such documents, provides a strong foundation for future research into this.

Elizabeth Chason (New Horizons Governor's School of Science and Technology).

Determining the Optimal Training Data Size for Recognizing Handwriting.

Abstract. This project sought to examine how different amounts of training and testing data influenced the precision and recall when recognizing handwriting through machine learning. Using the programming language Python, this project tested the computer's capacity to refine itself and create rules to predict what handwritten digit it was given from the MNIST data set of images of handwritten numbers from 0-9. Three different training data sizes, 70%, 80%, and roughly 86% of the total data, were used when testing a program for precision and recall. The optimal data set size for training the classifier was predicted to be 80% of the total data set, with the other 20% saved for testing data. A confusion matrix, showing the prediction behavior of the program, was produced and from it the F_1 score was calculated. The training data size with the highest F_1 score was deemed the most efficient. The study showed that 70% yielded an average F_1 score of 0.903, 80% yielded 0.901, and 86% yielded 0.899. The smallest training set performed the best according to the study's standards, as it had the F_1 score closest to 1. Machine learning has been revolutionary in industries such as the postal service, as now computers can read addresses and zip codes with incredible speed. However, by altering the amount of training data given to the computer, it is possible that even higher levels of efficiency can be reached in reading human writing.

Sophia Cronin (James Madison University), Tyler Webster (James Madison University) and Jason Forsyth (James Madison University).

Student Research Short: Wearable Computing for Physical Rehabilitation.

Abstract. According to the National Library of Medicine, more than 50% of the US population is affected by musculoskeletal impairments, making it the nation's leading cause of disability. When treating this, a patient undergoes Physical Rehabilitation, and is instructed to perform exercises at home, but they often forget to perform them or do so incorrectly which hinders their recovery.

The proposed research seeks to use wearable computing technology to create a system that detects exercise performance, sends collected data to the physician, and provides immediate feedback to the patient. The team is exploring how they can provide effective client feedback to increase patient performance and self efficiency. Currently, the team is exploring light and haptic feedback which have shown high success rates in similar settings. Considering haptic feedback, the team has implemented vibration motors to send "push" signals that mimic the guidance of a physical therapist pushing patient motions into correct positioning. Considering visual feedback, the team implemented a circular arrangement of lights that guide the patient in the direction associated with where on the circle the light is illuminated.

To test ideas, we have conducted pilot studies where users performed simple predetermined exercises while receiving visual or haptic feedback from prototypes. Next, we plan to conduct a study where we can test a combination of both using a combined prototype. Ultimately, a final product would consist of a wearable device that can read a patient's data to provide immediate feedback to the patient and summarized feedback to the physician.

Natalia Ermicioi (Marymount University).

Factors affecting nonprofits' information security readiness during crises: A study of COVID-19 impact on Small and Medium Nonprofit Organizations (NPOs) in DMV Area.

Abstract. Considering the unprecedented COVID-19 pandemic crises, many NPOs, like other types of institutions, have migrated their operations remotely to maintain at least a minimum level of activity. Consequently, the NPOs are experiencing an increased level of online activity during Covid-19. This study examined the emerging phenomenon and any ensued cybersecurity threats and attacks by determining whether any relationship exists between the number of employees, budget size, and beneficiaries of an NPO (independent variables) and the level of information security measures taken (dependent variables) during the Covid-19 pandemic crises. The findings of this research enlarge the limited body of current knowledge regarding NPOs' cybersecurity practices.

Brandi Fabel (Ivy Tech Community College).

Cybersecurity Considerations in Small Business, Economic Impact.

Abstract. The cyber industry is a combative market. It is called cyber warfare for a reason.

Reactive methodology in cybersecurity, as a leading response, creates this constant state of unrest.

Cybersecurity organizations such as the Open Web Application Security Project (OWASP) encourage security analysis and application within the Software Development Lifecycle (SDLC) throughout system

development (OWASP, 2018). This suggests the technology industry recognizes the need for notional security analysis, but that, “in practice, threat models are often created for existing systems, making it part of maintenance” - an afterthought to project implementation (Fahey, 2018).

We have opportunity to forge new methodologies because of cyber needs in unprecedented environments. This challenge, along with solutions it brings, can sharpen the horizon for our own local communities. Small businesses with fewer cyber-attacks means fewer negative financial impacts in local communities; thus providing relief on local economic posture, and holistically, the cyber economy.

This research examines security concepts, such as threat modeling, that work alongside the Software Development Life Cycle for technology projects of any size or purpose in small business environments. Further, this study concept specifically addresses how to categorize and organize threats as they uniquely map to the business requirements; an approach that allows small business to improve security postures using existing frameworks and guidance as a template.

The presentation will show the "how" to connect and address these requirements, and the potential it has in small business environments.

Jennifer Hurst (Marymount University) and Dr. Faleh Alshameri (Marymount University).

The Dark Side: A dive into Dark Patterns and Data Brokers.

Abstract. It is very distressing that consumers often allow their personal private information to be collected and shared so freely these days. They do not seem to be aware of what information is being collected, or when and who can access it. In order to use some apps and services, the consumer is encouraged to provide more information than is needed, and this is a deceptive practice called a Dark Pattern.

This study explores various Dark Patterns and Data Brokers. It recommends steps to protect consumer privacy, curtail financial harm, and prevent consumers from providing information a merchant does not need in order to obtain their services.

We collected and analyzed the opt out information from 14 Data Broker websites that sell consumer data for marketing and other purposes. Then we analyzed the content and developed categories to determine if a Dark Pattern was present, which one if so, and how easy or difficult it was for a consumer to actually complete the opt out in the website.

The potential outcome of this study is that there should probably be regulation in the United States about what can and can’t be done to direct consumer behavior, for the collection and use of their information and to regulate how consumers can opt out to include making it obvious and easy.

Lalitha Kuppa (Virginia Tech), Tanvi Haldankar (Virginia Tech) and Margaret Ellis (Virginia Tech).

A Walk Down Memory Lane: Analysis of Memory and Computer Systems From 1995 to Present-Day.

Abstract. Since the dawn of computing, the world has tracked system performance. Yet, computer system performance data is still primarily siloed by benchmark, system, or system component. The mission of the Computer Systems Genome Project (CSGenome) is to draw together this data to analyze the evolution of system architecture and performance. This work aims to discuss how memory in

particular has evolved, both theoretically and in practice from the mid-1990s to present-day, and how that has impacted overall performance.

Firstly, the growth of memory was analyzed in a broader context by studying its development in major computing systems over the past three decades. In particular, different generations of synchronous dynamic random access memory (SDRAM) technologies, Rambus DRAM, and several asynchronous DRAM modules were investigated.

To quantitatively assess the evolution of memory, data from the CSGenome repository, previously scraped from the Standard Performance Evaluation Corporation (SPEC) Benchmarking Suite, was standardized to extract memory attributes, such as RAM generation, bandwidth, and latency. Benchmark performance was also standardized over the four generations of SPEC Benchmarks (CPU95, CPU2000, CPU2006, and CPU2017) to determine trends across systems. Beyond this, the accuracy of results was validated by analyzing data against known memory trends.

The addition of standardized memory data to the CSGenome repository provides a way to observe trends across multiple components in computer systems. As CPU-intensive benchmarks, such as the SPEC Suite, generally emphasize a system's processor, memory subsystem, and compiler, this repository provides a way to isolate trends related to individual system components.

Jing Ma (University of Virginia).

Multi-Cause Effect Estimation with Disentangled Confounder Representation.

Abstract. One fundamental problem in causality learning is to estimate the causal effects of one or multiple causes (a.k.a. treatments, e.g., medicines in the prescription) on an important outcome (e.g., cure of a disease) for each individual (e.g., a patient). One major challenge of causal effect estimation is the existence of unobserved confounders -- the unobserved variables that affect both the treatments and the outcome. Unobserved confounders may bring strong biases to causal effect estimation. Recent studies have shown that by modeling how individuals are assigned with different treatments together, the patterns of unobserved confounders can be better captured, and thus the estimation biases can be eliminated.

However, the interpretability of the captured unobserved confounders in these works is limited, i.e., it is difficult to understand the semantic meaning of the captured unobserved confounders, as they are often mixed together. To address this problem, we focus on the multi-cause effect estimation problem from a new perspective by learning disentangled representations of unobserved confounders. The disentangled representations are vectors which include information to describe the confounders, and encourage different parts of the vectors to describe different confounders. In this way, the disentangled representations of confounders can not only facilitate the causal effect estimation but also strengthen the understanding of causality learning process. Experimental results on both synthetic and real-world datasets show the superiority of our proposed framework from different aspects.

Emma Macaluso (James Madison University) and Kevin Molloy (James Madison University).

Characterizing Antimicrobial Peptides with Computational Techniques.

Abstract. In 2019, the CDC reported that 35,000 people died from bacterial infections due to antibiotic resistant strains. Research into antimicrobial peptides (AMPs) shows a promising new alternative to failing antibiotics. AMPs are naturally occurring proteins that act in a host's innate immune response to bacteria. The mechanisms AMPs use to kill bacteria are an ongoing area of research. Most studies

have focused on an AMPs amino acid sequence. Structural information is more challenging to determine because AMPs are flexible molecules. However, structural information has the potential to be even more valuable than sequence information. Structure is the key to understanding the function of a protein. Because structure is so important to proteins, this work focuses on utilizing structure to perform the task of AMP recognition, that is, whether or not a given peptide exhibits antimicrobial properties.

This work hopes to enhance our understanding of the role structure plays in how AMPs perform their biological function. To this end, this work establishes a public labeled benchmark dataset for this and future AMP research. These structural datasets are valuable to the AMP community because most, if not all, AMP datasets consist only of sequence data. This work also utilizes the new structural data to augment and improve AMP recognition. This ongoing work is exploring different machine learning network designs and comparing their performances.

Darshini Mahendran (Virginia Commonwealth University) and Bridget McInnes (Virginia common wealth university).

RelEx: A system for multi-class clinical relation extraction.

Abstract. Relation extraction is a natural language processing (NLP) to detect and classify the relation between two entities in a text. Due to the exponential growth of text in recent years, automatic extraction of semantic relations from text has received growing attention. In this work, we explore three deep learning-based approaches for the multi-class classification of relations. The first explores three Convolutional Neural Networks (CNNs) architectures; one being a novel multi-label architecture. The second utilizes Bidirectional Encoders Representation from Transformers (BERT) language models. The third proposes a hierarchical based approach to remove the influence of the negative instances during the multi-class classification. We evaluate our method on a clinical dataset annotated for medical problems, treatments, and tests; and their relations. We report the precision, recall, and F_1 scores and compare our method to six current state-of-the-art approaches. Our results show that our novel multi-label CNN architecture obtained a higher F_1 score overall and outperforms the other CNN architectures; classes with fewer instances perform better with BERT-based models, and there is a significant improvement in the performance across all three CNN models when applying the hierarchical based approach.

Brooke Nixon (Christopher Newport University).

The Impact of Mobile Technology on Child Development.

Abstract. While computer science has brought rapid developments to mobile technology, it is essential to understand the benefits and detriments of these new devices so that negative implications can be avoided. The impact of technology use on child development is a question that has attracted substantial theorizing in the past decade as smartphones and tablets have become increasingly pervasive in usage. Recent statistics revealed that one third of preschool aged children in the United States now own and regularly use a tablet, but there is a lack of data to show the impacts of this. The preschool age is also when critical changes in theory of mind (ToM) understanding, or the ability to attribute mental states to one's self and others, occur. ToM acquisition is a fundamental part of child development, and a deficient ToM can have substantial impacts. This flash talk will examine the connections and impact between technology use and ToM development through looking at a research

study being conducted in Summer, 2021. In my talk, I will explain the need for research on such impacts within the computer science field and the benefits of technology as a mode of presentation in certain types of psychology testing. How technology has impacted psychological research, why it is more engaging to children, and how to apply this information in the computer science field will also be explored. By synthesizing these two fields, this talk will provide information about a frequently unarticulated yet exceedingly relevant aspect of the computer science field.

Shiva Omrani Sabbaghi (George Washington University) and Aylin Caliskan (George Washington University).

Measuring Gender Bias in Word Embeddings of Gendered Languages Requires Removing the Grammatical Gender from the Vector Space.

Abstract. Numerous studies have shown that machine learning models can learn human-like biases and perpetuate them in the society. Word embeddings, which learn word vector representations based on distributional semantics, are one such example. Measuring bias in word embeddings derived from grammatically gendered languages requires taking the language structure into account. Otherwise, grammatical gender signals lead to inaccurate and stereotype-incongruent results where ground truth social bias information is available. Grammatically gendered languages assign gender to nouns and enforce accompanying words to agree with the gender. Word embeddings learn grammatical gender signals via the association between nouns and their gender, which outweigh social gender bias in magnitude. As a result, bias quantification methods like the Word Embedding Association Test (WEAT) need to first remove grammatical gender information to measure social gender bias in gendered languages. For the first time, we generalize WEAT from English to six grammatically gendered languages from the Germanic,

Romance, and Slavic branches of the Indo-European language family. We show that applying WEAT in these languages without removing grammatical gender may lead to stereotype-incongruent results, especially when measuring gender bias. Consequently, we introduce methods to identify and evaluate grammatical gender signals in word embeddings. After removing grammatical gender, we apply WEAT to demonstrate that the computed bias scores match the aggregate stereotype-congruent ground truth scores reported by millions of human subjects in social psychology who are native speakers of these languages.

Maya Ramani (University of Virginia), Aaron Thompson (University of Virginia) and John Hott (University of Virginia).

Can Patterns in Mystery Novels Be Detected?

Abstract. This research involves the analysis of mystery novels and stories using Natural Language Processing (NLP) and network analysis to accurately determine the identity of the guilty suspect. We are creating a “detective algorithm” that combines various NLP computing techniques to parse and process the narrative, and output the predicted identity of the guilty suspect along with a prediction score. First, the sentiment analysis of the text surrounding each character mention was collected to see how that character is discussed by the author and other characters and to ascertain if the context surrounding the mention of their name may have a unique pattern or connotation. Next, network analysis was used to provide a representation of the interconnectedness of the characters in the story arc, based on how many contexts they were involved together in, whether that be a mention or real time conversation. Network patterns across narratives are compared to identify patterns in how the

guilty suspect is interconnected with the other characters in the story. We are currently investigating our algorithm on Sir Arthur Conan Doyle's Return of Sherlock Holmes collection and hope to apply it to deduce the guilty party in other stories. Our project helps strengthen the bridge between the fields of computing and literature, as it explores the capabilities of an algorithm to be fine-tuned to predict human behavior in stories. Creating a "detective algorithm" also provides more insight into how predictable the mystery genre is, creating a new lens through which literature can be studied.

Autumn Toney (George Washington University) and Aylin Caliskan (George Washington University).

A Novel Word Embedding Intrinsic Evaluation Test: Quantifying Valence in Semantics Derived from Language Corpora.

Abstract. Word Embeddings are vector space representations of words learned by a machine learning model from a given text corpus. These word vectors capture the lexical semantics in the given corpus, proving to be effective in various Natural Language Processing (NLP) tasks. Traditionally, the quality of word embeddings is measured by intrinsic evaluation tasks, which analyze and interpret the syntactic and semantic characteristics of word embeddings (e.g., word similarity), or extrinsic evaluation tasks, which measure how well word embeddings perform on downstream tasks (e.g., part-of-speech tagging). Currently, the most common intrinsic evaluation tests are word similarity and word analogy tasks.

Extending methods that quantify human-like biases in word embeddings, we introduce ValNorm, a novel intrinsic evaluation task and method to quantify the valence dimension of affect in human-rated word sets from social psychology. We apply ValNorm on static word embeddings from seven languages (Chinese, English, German, Polish, Portuguese, Spanish, and Turkish) and from historical English text spanning 200 years. ValNorm achieves consistently high accuracy on quantifying the valence of non-discriminatory, non-social group word sets. Specifically, ValNorm achieves a Pearson correlation of $p = 0.88$ for human judgment scores of valence for 399 words collected to establish pleasantness norms in English. In contrast, we measure gender stereotypes using the same set of word embeddings and find that social biases vary across languages. Our results indicate that valence associations of non-discriminatory non-social group words represent widely-shared associations, in seven languages and over 200 years.

Ana Valentin (Marymount University).

Exploring technical skills and competencies for future information technology workforce.

Abstract. The purpose of the study is to explore the relationship between the top five information technology (IT) skills needs and employees' IT competencies to strengthen workforce capabilities toward a competitive global business market. In Phase 1, 105 out of 453 (23.2%) supervisors identified the top five IT skills needs as big data, cloud, program, and planning management, cybersecurity/security, artificial intelligence, and machine learning. Although on Phase 2, 376 out of 878 (42.8%) employees self-reported that their level of IT competencies as Expert was the user and customer support, systems administration, administration and management, information assurance, and information technology architecture. The analysis of variance concluded that there was a relationship between IT skills needs and the level of IT competencies of the employees for a P-value

less than 0.05. The results of the study recommended the modernization of the organization's workforce plan to reduce gaps of IT skills needs and competencies among existing employees.

Vanamala Venkataswamy (University of Virginia).

Renewable Energy Aware Job Scheduling in Green Data Centers.

Abstract. The exponential growth in demand for digital services is driving massive data center energy consumption and negative environmental impacts. It is crucial to promote sustainable solutions to pressing energy and digital infrastructure challenges. The hyperscale cloud providers are moving towards data centers powered by renewables. Co-locating the data centers near green energy production sources further reduces energy transmission loss and increases the energy available to the data centers. However, integrating wind and solar to power the data centers is challenging because the power generation is intermittent. The data centers can employ multiple renewable energy sources (solar and wind) by intelligently adapting computing to energy generation using smart systems and system-software. We present the Leaf, a Deep Reinforcement Learning (DRL) based job scheduler that automatically learns efficient job scheduling policies while continually adapting to data centers' complex dynamic environment.

Lijing Wang (University of Virginia).

Using Mobility Data to Understand and Forecast COVID19 Dynamics.

Abstract. Disease dynamics, human mobility, and public policies co-evolve during a pandemic such as COVID-19. Understanding dynamic human mobility changes and spatial interaction patterns are crucial for understanding and forecasting COVID19 dynamics. We introduce a novel graph-based neural network(GNN) to incorporate global aggregated mobility flows for a better understanding of the impact of human mobility on COVID-19 dynamics as well as better forecasting of disease dynamics. We propose a recurrent message passing graph neural network that embeds spatio-temporal disease dynamics and human mobility dynamics

for daily state-level new confirmed cases forecasting. This work represents one of the early papers on the use of GNNs to forecast COVID-19 incidence dynamics and our methods are competitive to existing methods. We show that the spatial and temporal dynamic mobility graph leveraged by the graph neural network enables better long-term forecasting performance compared to baselines.

Susan Zehra (Old Dominion University), Syed Rizvi (Old Dominion University) and Stephan Olariu (Old Dominion University).

Auction-oriented Parking-based Vehicular Cloud Management.

Abstract. Researchers have shown that most vehicles are parked for the majority of their time at parking garages, parking lots, or driveways. When parked, the computing resources of these vehicles are unused and untapped. Vehicular Cloud is currently an area of research that has attracted substantial attention, primarily due to the success of conventional cloud services. In this model, each vehicle is a computation node. The availability of nodes is the main discriminant between traditional cloud computing and vehicular cloud computing. Unlike traditional cloud, the nodes are not available all the time in vehicular cloud. Vehicles create a dynamic environment as they randomly enter and exit parking garages. Therefore, node (vehicles) are only available during the time when vehicles remain

parked in parking lots. In this paper, we introduce a novel auction-based framework for Vehicular Cloud. It is based on auction and market design approaches. The major contributions in this paper are: (1) a concept of how to integrate sophisticated software agents that can search, bid, price, and allocate jobs on behalf of the stakeholders, (2) formulation of a auction-based job management that unifies the strategies related to job allocation, scheduling, and pricing, and (3) simulation studies to demonstrate that our economics achieve substantial performance benefits. Simulation results show that the proposed interactive agents in our model enable efficient processing of a vast amount of data, providing cost savings to the stakeholders, ultimately lowering the load from conventional cloud, and improving the utility of parked vehicles and parking facilities.

Technical Workshops

Leslie Fiorenzo (Coach With Leslie, LLC).

Difficult Conversations – 3 Keys to Mastery.

Abstract. Technical professionals are frequently involved in difficult conversations. We all avoid conversations that are conflictual or uncomfortable. Often, those are exactly the conversations we need to have. Are you struggling to discuss an important issue with a colleague, customer, direct report, or your supervisor? Then this session is for you. You will leave with a road map and tools in your tool- box to use as you navigate the conversation you've been avoiding.

This session is designed to give technical professionals 3 tools to immediately take back to their workplace to use and share with other leaders. They will be able to create a structure that is easy to follow and implement, identify the words to get the conversation off to a positive start and how to create allies instead of an enemies.

Session objectives

A proven formula – BECAUSE – to create a framework for any difficult conversation. Being able to understand the components of what works allows each participant to confidently manage any difficult conversation.

Words matter. In this presentation, participants will learn specific language to set up any conversation for success. Participants will also learn the formula for creating their language and approach.

Participants will also learn how to drive understanding and consensus to achieve buy in from the other person as well as creating a strong, lasting alliance.

Peter Henderson (Butler University, Emeritus).

How do Computer Systems Work?

Abstract. I have been giving presentations in local schools on various computing topics. One fifth grade teacher asked me to explain to her class how computers work. I will sum up my in-depth research with the quote "How computers work is very complex" from one educational blog.

In response, I developed a presentation, using a variety of kinesthetic activities, which convey a basic understanding of how a simple Central Processing Unit and Memory work to run simple instructions for input, output, data movement and basic arithmetic. Due to Covid19 I have not used this. However, I have extended the central ideas to a wider audience, and have done several virtual middle school technology presentations, and several virtual Osher Lifelong Learning classes at William & Mary. For the latter, which was a very diverse audience, in a 2 hour interactive session, the participants gained a basic understanding of the important concepts – based on course evaluations.

In this virtual workshop I will share my ideas, philosophy, and material, which includes powerpoint presentations, reinforcing worksheet activities, and in class kinesthetic activities.

Panels

Christina Burton (University of Pennsylvania, School of Engineering and Applied Science) and Rita Powell (University of Pennsylvania, School of Engineering and Applied Science).

Graduate School - Is It for Me?

Abstract. Are you interested in graduate school but not sure it is a fit for you? What will a Master's degree do to enhance your job prospects, both in the short and long run? Do you have specific interests that may not be addressed in a standard Master of Science in Engineering program? The 6 student panelists have a diversity of interests to share since they are all in different computer science graduate programs at the

University of Pennsylvania. They will tell you how they identified the right program for them, what led up to their selection of their graduate program and the dream jobs they hope to land following graduation.

Birds of a Feather

Weizhen Sheng (Percentage Project), Stephanie Tang (Percentage Project) and Lucia Huo (Percentage Project).

Campus Climates & Data-driven Advocacy.

Abstract. Do you feel intimidated studying Computer Science? Do you experience microaggressions? Do you have a professor/mentor whom you look up to? The Percentage Project, a 501(c)(3) nonprofit organization that focuses on data-driven advocacy, asks powerful questions like these in its annual climate survey for post-secondary school students.

Our 2019-20 survey found that 62% of students identifying as women (versus 27% men, 41% non-binary) agreed that they feel intimidated studying CS. We use this survey result alongside others to help inform the audience on the current climate of college computer science communities. We pair these statistics with an annual photography campaign on social media, presenting a "Humans of New York" approach to understanding the experience of computer science students today.

In this BOF session, we will briefly present select statistical results from our annual surveys that are most insightful or thought-provoking, as well as share some highlights from our past annual social media campaigns. We will use these as a springboard for participants to discuss and reflect on their own student experiences and to critically consider the climate at their school's computer science department.

Technical Talks

Gretchen Achenbach (National Center for Women & Information Technology and University of Virginia).

Discover the Power of Community with WIC Groups and NCWIT Aspirations in Computing.

Abstract. Groups for women in computing are important sources of community and support, but may risk creating the impression that women need "extra help" or divert attention from the societal conditions in which women's underrepresentation occurs. However, when viewed as one part of a multipronged strategy, Women in Computing (WIC) groups on college campuses and NCWIT's Aspirations in Computing can be powerful forces for change. Find out how to get involved in and leverage these communities to broaden participation in computing.

Shubhi Asthana (IBM Almaden Research Center), Bing Zhang (Texas A&M), Aly Megahed (IBM Almaden Research Center), Pawan Chowdhary (IBM) and Taiga Nakamura (IBM).

Demand Forecasting of Service Request (SR) Volume with Human-in-the-loop (HitL).

Abstract. Time-series forecasting is a well-explored topic in machine learning. However, in some forecasting applications, time-series models such as LSTM, ARIMA, VARMAX, etc. are not able to capture all relevant features and external factors that may influence prediction. Additionally, when dealing with high dimensionality data, the task of extracting important and discriminative features is non-trivial. We present an application in the services industry where we developed a time-series forecasting model with Prophet and augmented HitL to improve prediction. Our method leverages HitL to augment this time-series model and navigate with better features to achieve optimal performance.

We applied our novel method to an application of a global IT service provider, which has thousands of product portfolios that sell to enterprise clients. Due to varied range of products and services, a sales support person may not have complete knowledge of all products, and would require assistance from domain experts. In this case, they submit Service Request (SR) tickets to domain experts to get assistance.

However, in order to manage appropriate staffing levels in domain expert teams, one needs to know the volume of demand for SR tickets ahead of time.

Our proposed HitL method is comprised of four key steps:

- Exploratory data analysis and time-based feature engineering on historical data
- Recommendation engine which explores features to forecast the SR volume
- Hybrid Forecasting with HitL
- Allocation of workload for each SR ticket ahead of time.

Encouraging results were obtained on a real-world SR dataset with > 1M SR tickets.

Neli Chakarova (IBM).

Imbalanced data can throw you off balance.

Abstract. Imbalanced data can throw you off balance

The goal of the talk is to present the problems that might occur when dealing with imbalanced data. It is a common challenge and understanding how this affects the metrics that we use and how we can solve the problem could be very useful.

The initial agenda is the following:

0. Start with an example of a classification problem for which we have an imbalanced dataset.

1. Briefly explain some of the algorithms that we can use to solve the task as this will not be to focus of the talk – supervised machine learning algorithms like logistic regression or decision trees.

2. Introduce one of the most well-known metrics for evaluating how our model performed – accuracy. Then focus on why the very good results that we get for the accuracy are actually very deceiving and totally unreliable.

3. Introduce two other metrics that give us better understanding of the model performance. - precision and recall Explain how they are calculated mathematically and how they can be translated in business terms which helps us decide which one make sense to track.

4. Finally, I would like to mention a couple of techniques that could be used to solve the imbalanced data problem such as under-sampling, over-sampling, mixed techniques, synthetic minority over-sampling technique (SMOTE) and other and which python packages could be used for that.

Zhuojun Duan (James Madison University).

Game theory in Mobile Crowdsensing Networks (MCS).

Abstract. In mathematical economics and business, game theory studies the interactions among players, who take actions to influence each other and usually have conflicting or common benefits.

Recently, game theory is also widely used in computer science areas to balance the resources sharing among multiple mobile devices. In Mobile Crowdsensing Systems (MCSs), motivating mobile users to participate in sensing services for efficient data generation and collection is one of the most critical issues. In my presentation, I will introduce how to use game theory-based mechanisms to incentivize mobile users promisingly and effectively.

Agoritsa Polyzou (Georgetown University) and Lisa Singh (Georgetown University).

Considerations on Ethics, Machine Learning, and Social Media Research.

Abstract. Policymakers and researchers are demonstrating increasing interest in the ethical implications related to applied models and technologies. Based on their outputs and predictions, they drive decision-making on different aspects of our everyday lives. As they may positively or negatively affect their users, we need to ensure that their behavior aligns with some basic ethical principles. Specifically, in social media research, algorithms and machine learning models are developed to better understand human demographics, opinions, and behaviors. Over the past years, fairness and privacy are two areas studied in this context. While we have begun to grapple with some ethical issues associated with using social media data, there are still many unaddressed complexities related to the computational tasks we work on, the data we use, and the technologies we create. In this talk, we will

focus on the unique characteristics of social media data and the related ethical considerations that these characteristics create. We frame several ethical dilemmas and show how and when ethical concerns arise. While there are numerous aspects that we could consider, we will focus on a couple of examples and propose ways to quantify specific ethical violations. By cataloging these complexities and their ethical considerations, we take the first step towards addressing these issues as we continue to build algorithms and tools for social media.

Laura Puaca (Christopher Newport University).

#ilooklikeanengineer: Challenging Gender Stereotypes Since 1940.

Abstract. This talk will focus on early efforts to expand U.S. women's participation in engineering, beginning in World War II. In the two decades before the women's movements of the 1960s, a number of women and women's organizations (such as the Society of Women Engineers) tirelessly campaigned to enhance female representation in the field. Amidst the Second World War and Cold War, these reformers not only worked to gain women's admittance to previously all-male engineering programs and schools, but also collaborated with government officials, teachers, and guidance counselors to help reshape public perceptions of engineering as a career for women. In doing so, they established scholarships, published vocational guidance materials, spoke at career days, judged science fairs, and created mentoring opportunities for young women.

These early efforts are significant because they paved the ways for later efforts to encourage women in STEM. Throughout the 1970s, 1980s, and even today, proponents of women's engineering education and employment continue to rely on many of these strategies. They also face many of the same obstacles with regard to gender stereotyping. By looking at past and present efforts to improve women's engineering participation as well as representation of women engineers, this lecture will shed light on the gains achieved to date as well as the ongoing struggle to improve women's acceptance in the field.