

# CAPWIC 2022

The ACM Capital Region Celebration of Women in Computing

*Abstracts listed in alphabetical order by first author.*

## Contents

Keynotes.....	2
Panels.....	4
Research Shorts .....	6
Student Posters.....	11
Flash Talks .....	15
Birds of A Feather .....	20
Technical Workshops .....	21
Technical Talks .....	22
Sponsors / Career Fair.....	23



## Keynotes

Ria Galanos ([Yext, American University](#)).

### ***Transitioning from College to Industry.***

Ria Galanos is the software engineering lead for learning and development at Yext. She creates materials for new engineers at Yext to help them get acclimated to Yext's systems, tools, and workflow so they feel empowered to tackle their first tasks on their own. She tries out new ways to increase the diversity among the engineers at Yext and mentors both high school and college students to help retain them in the field. Ria noticed upon joining Yext that there isn't a great way to predict how a candidate is going to perform during technical interviews based on their resume. After poking around, she has learned that students have drastic differences in their familiarity with the software engineering interview process. She is working to change that so some students aren't at a disadvantage before they arrive for their interview.

Ria is also an adjunct professor in American University's Computer Science Department. She runs professional development workshops for AP Computer Science A teachers for the College Board and has participated in both the scoring of and writing of the AP Computer Science A exam. Prior to joining Yext, Ria taught high school computer science for 16 years. Her current educational passions are building the confidence of women in computer science courses and creating interdisciplinary real-world computer science assignments that help make computing relevant and personal for all her students. She was awarded the 2011 Aspirations in Computing Educator Award sponsored by the National Center for Women & Information Technology and Google for her support of young women's participation in computing and technology. Ria holds a BS in Aeronautical Engineering from Rensselaer Polytechnic Institute, a MAEd in Secondary Mathematics Education from Virginia Polytechnic Institute and State University, and an advanced teaching diploma from Emory University.

Colleen Lewis ([University of Illinois at Urbana-Champaign](#)).

### ***Imposter Syndrome.***

Hi! My name is Colleen - I'm a CS prof at the University of Illinois at Urbana-Champaign and I research how people learn CS and feel about learning CS. I'm excited to share some of my stories about my path in CS and life! For example, you probably wouldn't guess that when I first took data structures as an undergrad, I had to late drop the class so that I wouldn't fail. I was really lucky that none of my friends or family told me I just wasn't cut out for CS. However, lots of random people told me that I shouldn't do CS b/c I'm a "people person." What nonsense - there isn't just one way to be in CS!

Colleen Lewis is an Associate Professor of computer science (CS) at the University of Illinois Urbana-Champaign. Lewis was previously the McGregor-Girard Associate Professor of CS at Harvey Mudd College. At the University of California, Berkeley, Lewis completed a PhD in science and mathematics education, an MS in computer science, and a BS in electrical engineering and computer science. Her research seeks to identify and remove barriers to CS learning and understand and optimize CS learning. Lewis curates CSTEachingTips.org, a NSF-sponsored project for disseminating effective CS teaching practices. Lewis has received the NSF CAREER Award, the NCWIT.org Undergraduate Mentoring Award and the AnitaB.org Emerging Leader Award for her efforts to broaden participation in computing.

Angela Orebaugh ([University of Virginia](#)).

***Being a Woman in Cybersecurity.***

Angela Orebaugh is saving the world bit-by-bit. She is an Assistant Professor at the University of Virginia's Computer Science Department where her teaching and research is focused on keeping us safe and secure in the cyber world. She was the 2020 recipient of the prestigious Adelle F. Robertson Award for Excellence in Teaching. Prior to joining academia as a professor, Angela worked in industry for over 25 years providing cybersecurity expertise to clients such as the National Institute of Standards and Technology (NIST), the Department of Defense (DoD), intelligence agencies, small businesses, and start-ups. Her mission in academia is to leverage her industry expertise to add value and create exceptional learning experiences for students.

Angela completed her Ph.D. at George Mason University's Volgenau School of Engineering with published papers in the areas of behavioral biometrics, data mining, authorship analysis, and cyber forensics. She completed her Master of Science in Computer Science and her Bachelor of Business Administration at James Madison University. Angela is an internationally recognized author of several bestselling technology books, over thirty published articles, and co-author of seven NIST publications. She was honored in the 2019 book *Women Know Cyber* as one of the 100 Fascinating Females Fighting Cybercrime. Dr. Orebaugh was selected as one of *Information Magazine's* Security 7 in 2013 and was appointed as a Booz Allen Hamilton Fellow in 2012.

## Panels

Clare Arrington ([Rensselaer Polytechnic Institute](#)). Christina Russo ([MasterPeace Solutions](#)). Harshini Sri Ramulu ([George Washington University](#)). Sydney Pugh ([University of Pennsylvania](#)).

### ***Pursuing Graduate School in Computing.***

**Abstract.** Are you unsure if you want to attend graduate school? Not sure what it even means to go to graduate school or how it differs from undergraduate programs? Already in graduate school but want advice on how to succeed while there? Then this is the panel for you! In this panel you'll hear from four current graduate students in PhD and MS programs, to discuss their experiences and give advice on when/if to go to graduate school, and how to thrive once there. Our panelists include:

**Clare Arrington.** Clare is a current PhD student at RPI, after graduating with a computer science degree from University of Mary Washington in 2019.

**Christina Russo.** Christina is currently a Software Developer at MasterPeace Solutions while simultaneously earning her MS in Data Science from Loyola University Maryland, after earning her BS in computer science from Loyola University Maryland in 2019.

**Harshini Sri Ramulu.** Harshini is a current PhD student at George Washington University. She recently moved universities mid-PhD, when her advisor moved to GWU from Max Planck Institute for Security and Privacy, and has previously worked as a UX Researcher and Designer after earning an MS in Human Technology Interaction and BE in Computer science and Engineering.

Karen Jackson ([Apogee Strategic Partners](#)). Robin Moses ([IRS](#)). Megan Underwood ([Delta Air Lines](#)). Dr. Lauren Vega ([Booze Allen Hamilton](#)).

### ***Leadership in Computing.***

**Abstract.** What does it mean to be a leader in computing, and how does one become a leader? Whether you are ready to step into a leadership role yourself, or just curious about what these women do in their leadership positions or how they got there, this panel is for you! In this panel you'll hear from four women who are leading in four different types of roles. They'll share how they got to where they are today, what it means to be a leader and in the type of leadership role they are currently in, and answer your questions about leadership in computing. Our panelists include:

**Karen Jackson,** President of Apogee Strategic Partners.

**Robin Moses,** Deputy ACIO of Enterprise Operations at the IRS.

**Megan Underwood,** Manager of Information Security Risk at Delta Air Lines.

**Dr. Laurian Vega,** Senior Technical Lead at Booze Allen Hamilton.

Rachel Shorey ([The New York Times](#)). Rachel Kemper ([RDA Corporation](#)). Dr. Nadine Shillingford ([Stage 2 Security](#)). Dolapo Martins ([Microsoft](#)).

### ***Careers in Computing.***

**Abstract.** Are you interested in hearing about many different careers you could consider after earning a computing degree? Perhaps unsure what you want to do, or what various job titles even mean? Then this panel is for you! Join four women who are working in various aspects of computing to learn about what they do, how they got there, and what you should do to prepare for similar careers. This panel will also discuss the many different types of careers open to people with a computing degree. Our panelists include:

**Rachel Shorey.** Rachel has an MS in Computer Science and undergraduate degrees in mathematics and linguistics and is currently a Senior Software Engineer at The New York Times, working on newsroom software and data projects.

**Rachel Kemper.** Rachel has an interdisciplinary BA in computer science and psychology and works as a Senior Software Engineer at the consulting company RDA Corporation, with extensive experience with web technologies.

**Dr. Nadine Shillingford.** Nadine has a BS, MS, and PhD in Computer Science. Previously a professor, she is now a Project Manager at Stage 2 Security, and has extensive knowledge in cybersecurity.

**Dolapo Martins.** Dolapo has a BS in computer science, BA in Classical Languages, and BS in Linguistics. She works as a Software Engineer at Microsoft and researches natural language technologies and low-resource languages as a graduate student at the University of Washington.

**Sydney Pugh.** Sydney is a current PhD student at University of Pennsylvania, after graduating with a BS in Applied Math and a BS in Computer Science from Loyola University Maryland in 2019. Flash Talks

## Research Shorts

Kimberlee Ann Brannock ([Marymount University and HP Inc.](#)).

### ***Cyber Security and Consumers.***

**Abstract.** We are all consumers, even those of us that are cyber security practitioners, cyber security professionals and cyber security experts. Current approaches to consumer cyber security have an emphasis in securing only one point that can be a significant cyber security risk or a few points of vulnerability, and do not do an adequate job of addressing overall cyber security risk in structured manner. As a result, a structured consumer cyber security solution needs to take into account behaviors, readiness, education, mitigation, and so forth. This is the approach that is already used by enterprises and networks, which includes the NIST cyber security framework. We need to approach cyber security end to end and have a cyber security framework that helps accomplish this goal. A structured consumer cyber security solution needs to take into account behaviors, readiness, education, mitigation, and so forth, which translates into a consumer cyber security framework.

Samantha Dies ([Georgetown University](#)) and Lisa Singh ([Georgetown University](#)).

### ***Quantifying and Measuring Transparency in Computer Science Research.***

**Abstract.** The exponential advances in the field of computer science, paired with increasing datafication, have led to a new environment in which research decisions and algorithms have direct impact on individuals' daily lives. Computer scientists are beginning to develop ways to measure ethical expectations of research. In this presentation, we focus on ethical expectations as they relate to transparency with respect to the data collection and preprocessing stages of the data science life cycle. We focus on computer science research conducted about mental health that uses social media data. We develop a rubric for expected transparency in relevant computer science research using eighteen criteria spanning research design explanation, data sample and ground truth discussion, human subject protection, and preprocessing justification. We find that current computer science publications studying mental health using social media have transparency scores between 31.25-100%, with an average score of 63%. We then demonstrate that we can build a semi-supervised model for determining transparency. Using a combination of clustering, topic modeling, and binary classifiers, we generate transparency scores. Our work is ongoing as we attempt to improve the performance of these models.

Alexis Kochanski ([University of Mary Washington](#)) and Veronica Cagle ([University of Mary Washington](#)).

### ***Using Machine Learning to Identify Political Polarization on Social Media.***

**Abstract.** Polarization in the political sphere, seen through combative communication and stalemate, may impose negative social impacts on the population. Attempting to measure political polarization in the masses through self-reported surveys and interviews can present response biases of social desirability. The classification of thought freely written online allows political polarization to be measured in an impartial manner. Reddit is one application that enables users to share opinions and create discussions anonymously; this text can be used to measure the political climate at any given time. Disagreement has grown over the perceived level of polarization in our society. The purpose of our research is to measure the degree of political polarization over time by collecting and classifying threads of dialogue within political communities on Reddit.

Our team utilized Reddit APIs to gather threads of text from multiple subreddit communities online. We recruited a team of evaluators to hand-annotate the threads for polarization. Then, we created a machine learning classifier to predict whether a thread posted on Reddit is polarized or not. We trained a multi-layer feed-forward neural network on the hand-tagged data. A neural network takes the data as input and goes through a series of layers to search for patterns in the text. To produce a more extensive training set out of our unlabeled data, we employed the method of bootstrapping. Preliminary efforts to bootstrap data have shown improvements in our classifier's accuracy and assisted in visualizing patterns of polarization over time.

George Li ([University of Maryland](#)), Ann Li ([University of Virginia](#)), Madhav Marathe ([University of Virginia](#)), Aravind Srinivasan ([University of Maryland](#)), Leonidas Tsepenekas ([University of Maryland](#)) and Anil Kumar Vullikanti ([University of Virginia](#)).

### ***Deploying Vaccine Distribution Sites for Improved Accessibility and Equity to Support Pandemic Response.***

**Abstract.** In response to COVID-19, many countries have mandated social distancing and banned large group gatherings in order to slow down the spread of SARS-CoV-2. These social interventions along with vaccines remain the best way forward to reduce the spread of SARS CoV-2. In order to increase vaccine accessibility, states such as Virginia have deployed mobile vaccination centers to distribute vaccines across the state. When choosing where to place these sites, there are two important factors to take into account: accessibility and equity. We formulate a combinatorial problem that captures these factors and then develop efficient algorithms with theoretical guarantees on both of these aspects. Furthermore, we study the inherent hardness of the problem, and demonstrate strong impossibility results. Finally, we run computational experiments on real-world data to show the efficacy of our methods.

Aashni Manroa ([College of William and Mary](#)).

### ***Modeling and Preventing Virus Spread.***

**Abstract.** This agent based model shows how to reduce the spread of a virus through preventative measures, which can be applied to the current COVID-19 pandemic. Agent based modeling is useful for this topic because interactions between individual people and the individual decision making process is difficult to represent without an ABM. This virus spreading model is unique because it is based on a geospatial element. Other COVID-19 models take into account hundreds of factors, but they do not include a geospatial element that actually shows the agents as they are moving and interacting with each other. Through various experiments, I have found that using both masks and vaccines as preventative measures are more effective in reducing deaths, while just masks as a preventative measure works best in reducing virus spread.

Elizabeth Miller ([Towson University](#)), Rashedur Rahman ([Towson University](#)), Moinul Hossain ([Towson University](#)) and Aisha Ali-Gombe ([Towson University](#)).

### ***What You Don't Know Might Hurt You.***

**Abstract.** Data privacy, a critical human right, is gaining importance as new technologies are developed and the old ones are evolving. In mobile platforms such as Android, privacy policy statements are used as a tool by app developers to communicate data access request. This case study cross-examines the privacy policy statements of popular social media apps, specifically Facebook and Twitter and quantifies the number of privacy policy statements that use ambiguity, sensitive data types, or do not request user permission through the Android permission model. Vague privacy policy statements are classified based on the requested data type or the purpose for data access. Subsequently, we conducted a comparative analysis examining trends in privacy statements that may constitute a threat to user security.

Brooke Nixon ([Christopher Newport University](#)) and Nicole Guajardo ([Christopher Newport University](#)).

### ***The Psychology of Texting: Mimicry in Computer-Mediated Communications.***

**Abstract.** The chameleon effect refers to humans' subconscious mimicry of the behaviors and mannerisms of others. The phenomenon is linked to one's desire to be liked by those around them and to their level of empathy. The chameleon effect has primarily been studied in-person, with recent research indicating the presence of the phenomenon in human-computer interactions. However, few studies have explored the chameleon effect in computer-mediated communications, specifically text messaging. This research short will explore a recent study on the chameleon effect applied to texting. In the study, participants viewed text message screenshots with either mimicked or opposing linguistic styles between the senders, then rated how likable the person replying to messages was. Results indicated the presence of the chameleon effect in computer-mediated communications with an empathy correlation and a gender effect. This talk will explore the background and results of the study, as well as the implications for such work, including how social mimicry applies to digital communications and chatbots.



Anooshka Pendyal ([Deep Run High School](#)).

### ***Comparing the Performance of Different Sized Convolutional Neural Networks on Cancer Identification.***

**Abstract.** Invasive ductal carcinoma (IDC) is the most common subtype of breast cancer, and it is important to identify it as quickly and accurately as possible. Currently, identification of IDC is done by hand, which is time-consuming and challenging. Machine learning, specifically convolutional neural networks, can be applied to assist in the diagnosis process. Currently, there is limited information about the effect of the number of layers a ResNet, a type of convolution neural network, has on the model's performance in cancer identification. In this project, the dataset was obtained from Kaggle, downsampled to address class imbalance, and preprocessed to prevent overfitting. Then, two models using different ResNets (ResNet34 and ResNet50) were trained and evaluated. One surprising result was that ResNet34 performed better during testing (F1 score of 0.71), while ResNet50 performed better during training (F1 score of 0.75) and validation (F1 score of 0.75). Generally, the deeper the neural network is, the more powerful it is, and the better it performs. Another result was that the validation accuracy and loss for both models was very unstable. This could be the result of insufficient training. Although the results of this project are inconclusive and the models do not perform well enough for real world use, there is still great potential for real world application, especially in assisting in the cancer identification process. It is important to study pre-trained models, as they reduce the amount of data needed to train a model, and data is difficult to access in healthcare.

Rebecca Rohan ([Marymount University](#)).

### ***Identifying Commonalities of Cyberattacks Against the Maritime Transportation System.***

**Abstract.** The purpose of this study is to identify commonalities in cyberattacks against the civilian maritime transportation system (MTS). The researcher conducted an exploratory document analysis to identify more comprehensive information and trends about the cyberattacks impacting and responsible adversaries conducting cyberattacks targeting maritime operations. The MTS can use any identified trends to make informed decisions about information technology (IT) and operational technology (OT) requiring new or enhanced cybersecurity measures. Current research identifies the trend of increasing cyberattacks against the MTS and examines publicly disclosed cyberattacks impacting MTS companies. However, current research fails to provide details about cyberattacks such as the tactics, techniques, and procedures (TTPs) employed or the impacted aspect(s) of the information security—confidentiality, integrity, and availability (CIA)—triad. Furthermore, current research on MTS cyberattacks fails to provide an in-depth examination of cyber adversaries, including social-political motivations or operator status (self-employed or contracted). At the conclusion of this limited, exploratory document analysis, the researcher determined the most targeted MTS aspect was shipping companies, followed by ports, administrative, shipbuilding, and vessel. Concerning the cyber adversaries behind MTS cyberattacks, most cyberattacks were conducted by China, followed by unknown cyber adversaries, Russia, Iran, and Israel. The most compromised aspect of the information security triad was availability, then confidentiality with no integrity compromises. Lastly, in terms of social-political categories, data exfiltration occurred the most, followed by ransomware, political agenda, and unknown.

Lynn Tao ([TJHSST](#)).

### ***A Novel Framework for Breast Cancer Risk Detection Leveraging Polygenic Risk Scores and Machine Learning.***

**Abstract.** Breast cancer (BC) is the most frequent cancer and the second leading cause of cancer death in women; an estimated 1 in 8 women in the US will develop breast cancer during her lifetime. However, current methods of breast cancer screening, including clinical breast exams, mammograms, biopsies, etc., are often under-utilized due to limited access, expense, and a lack of risk awareness, causing 30% (up to 80% in developing countries) of breast cancer patients to miss the precious early detection phase. This study creates a key step that the current BC diagnostic pipeline lacks: a pre-screening platform, prior to traditional detection and diagnostic steps. Our project presents BRECARDA, a novel framework that personalizes breast cancer risk assessment using AI neural networks to incorporate relevant genetic and non-genetic risk factors. A polygenic risk score (PRS) was enhanced by employing AnnoPred and validated by K-fold cross-validation, outperforming three current state-of-the-art PRS methods. The UK BioBank, including data from 170,417 women participants, was used to train the robust algorithm. Utilizing the enhanced PRS along with non-genetic information, BRECARDA was validated by 56,806 real-world female individuals and achieved a high accuracy of 93%, indicating its potential for clinical use. BRECARDA can alert high-risk individuals to get immediate screening, enhance disease risk prediction and diagnostic refinement, and improve the efficiency of population-level screening. Furthermore, BRECARDA can serve as a valuable platform to assist doctors in diagnosis and evaluation. BRECARDA is accessible, affordable, and efficient, and can save millions of lives.

Susan Zehra ([Old Dominion University](#)) and Syed Rizvi ([Old Dominion University](#)).

### ***A Novel Approach to Secure Vehicular Networks.***

**Abstract.** Vehicular networks have emerged as an effective technology for providing a wide range of safety applications for motorists and passengers. Today, vehicular networks have gained a lot of attention as they enable accessing a large variety of ubiquitous services. With popularity, the number of security vulnerabilities in their inter-vehicular services and communications have also increased. Consequently, the number of security attacks and threats have increased. It is of great importance to ensure the security of vehicular networks because their deployment must not compromise the safety and privacy of the stakeholders. To successfully defend against a wide range of attacks requires developing efficient and reliable security solutions and services. In this scientific research, a structured and novel approach is taken to provide security in vehicular networks. This study provides a new algorithm that secures a vehicular network and provides how it would defend when faced with vehicular security threats, vulnerabilities, and privacy issues.

## Student Posters

Eleni Adam ([Old Dominion University](#)), Tieming Ji ([University of Missouri at Columbia & Roche Diagnostics](#)) and Desh Ranjan ([Old Dominion University](#)).

### ***hummingbird: A software package for the detection of differential methylation.***

**Abstract.** Genes which are short sequences of DNA hold the information needed for all biological functions. Researchers have focused on identifying the relationship between the DNA and the observable characteristics of an organism. A vital mechanism that has been discovered to contribute in this is DNA methylation. Specifically, DNA methylation controls gene expression, meaning the activation of genes. The most common form of DNA methylation in vertebrates is the addition of methyl groups to specific Cytosine bases of DNA, prohibiting the corresponding genes in these locations from being expressed. In many cases of disease, like cancer, the methylation patterns differ between a patient and a healthy person. A region in the genome where mean methylation levels differ between two groups is called a differentially methylated region (DMR). We developed hummingbird, a package that enables the detection of such regions, the DMRs. The hummingbird uses a Bayesian Hidden Markov Model that incorporates location dependence among genomic loci, unlike most existing methods that assume independence among observations. It has three main functions: the hummingbirdEM, which executes the Expectation-Maximization algorithm for the Bayesian HMM and infers the best sequence of methylation states; the hummingbirdPostAdjustment, which allows researchers to place additional requirements on DMRs; and the hummingbirdGraph, which generates observation and prediction graphs for a region of interest. The hummingbird is faster and more accurate than the existing packages for detecting DMRs. For example, using as input data from the Large Offspring Syndrome study, the hummingbird generates the whole-chromosome 9 results in only 90 seconds.

Maytha Alshammari ([Old Dominion University](#)) and Jing He ([Old Dominion University](#)).

### ***Analysis of an Existing Method in Refinement of Protein Structure Predictions using Cryo-EM Images.***

**Abstract.** Protein structure prediction produces atomic models from its amino acid sequence. Three-dimensional structures are important for understanding the function mechanism of proteins. Knowing the structure of a given protein is crucial in drug development design of novel enzymes. AlphaFold2 is a protein structure prediction tool with good performance in recent CASP competitions. Phenix is a tool for determination of a protein structure from a high-resolution 3D molecular image. Recent development of Phenix shows that it is capable to refine predicted models from AlphaFold2, specifically the poorly predicted regions, by incorporating information from the 3D image of the protein. The goal of this project is to understand the strengths and weaknesses of the approach that combines Phenix and AlphaFold2 using broader data. This analysis may provide insights for enhancement of the approach.

Cassie Bedard ([James Madison University](#)), Garrett Hutson ([James Madison University](#)), Evelyn Munsterman ([James Madison University](#)) and Katelyn Anderson ([James Madison University](#)).

### ***Coffee Time Connections Using IoT.***

**Abstract.** Due to obstacles including lack of time or long distance, more people are routinely using the virtual environment to stay connected to friends and family. Our project's goal is to aid the older generation to utilize this environment to stay connected with their grandchildren. We are doing this by prototyping a coffee maker attachment, Coffee Time Connections, that enables users to send voice messages to each other and play them back while brewing coffee using the internet of things (IoT). We chose a coffee maker to implement our goal because making coffee is integrated into everyday life for all generations. The team designed Coffee Time Connections using universal platforms due to their low cost and extensive documentation. Buttons are connected to pins on a python programmed Raspberry Pi. The code downloads, plays, records, and uploads voice messages stored on Google Drive. We added two buttons (record and send) to the coffee maker which are placed in a 3-D printed box that stores all the added hardware. The record button allows the user to record a message, and the send button uploads the message to Google Drive for the paired coffee maker to access. The download and play functions are connected to the already existing brew buttons on the coffee maker. After a message is uploaded and sent, the corresponding coffee maker will download the message from Google Drive and play it aloud for the individual who pressed the brew button. This easy-to-use structure limits confusion and enhances connection between individuals.

Jing Ma ([University of Virginia](#)).

### ***Learning Fair Node Representations with Graph Counterfactual Fairness.***

**Abstract.** Fair machine learning aims to mitigate the biases of model predictions against certain subpopulations regarding sensitive attributes such as race and gender. One of existing fairness approaches, counterfactual fairness, compares the predictions of each individual with predictions of the same individual, but having modified their sensitive attribute (the modified versions of an individual are called its counterfactuals). Recently, a few works extend counterfactual fairness to graph data, but most of them neglect the following facts that can lead to biases: 1) the sensitive attributes of each node's neighbors may causally affect the prediction w.r.t. this node; 2) the sensitive attributes may causally affect other features and the graph structure. To tackle these issues, in this work, we propose a novel fairness notion – graph counterfactual fairness, which is a notion of counterfactual fairness on graphs, and it considers the biases led by the above facts. To learn node representations towards graph counterfactual fairness, we propose a novel framework based on counterfactual data augmentation. In this framework, we generate counterfactuals corresponding to perturbations on each node's and their neighbors' sensitive attributes. Then we enforce fairness by minimizing the discrepancy between the representations learned from the original graph and the counterfactuals for each node. Experiments on both synthetic and real-world graphs show that our framework outperforms the state-of-the-art baselines in graph counterfactual fairness, and also achieves comparable prediction performance.

Maya Narayanasamy ([University of Maryland, College Park](#)).

### ***Increasing Confidence and Interest in STEM: A 3-Year Computing Camp as a Gateway to Persistence for Young Women.***

**Abstract.** As the number of young women earning undergraduate degrees in computing fields in the United States remains disproportionately low, informal K-12 learning environments, such as summer camps, aim to combat this underrepresentation by increasing girls' confidence in STEM. CompSciConnect (CSC), a three-year summer computing camp for middle school students, exposes students to encouraging and supportive STEM near-peer mentors, and to an engaging STEM curriculum in an immersive environment. Factors such as early intervention and informal learning programs may be a key component to retaining young women on a longitudinal level. This study aims to investigate the impact that early intervention through informal computing programs like CSC have on young women's confidence over the course of their three years at camp. Using pre- and post- survey data across the years 2016 to 2021 and data from a longitudinal survey aimed at CSC alumni considering life after high school, initial findings indicate that summer computing camps positively impact student confidence in regards to whether they can succeed in a STEM field and do well in their science and math courses, increasing their likelihood of later persisting in STEM. This research highlights that informal computing camps aimed towards middle school students can be a critical addition to boosting young women's overall confidence in themselves and their computing abilities during their early adolescence.

Sharon Lee Reagan Díaz Bba ([Marymount University](#)), Khoanam Nguyen Msc ([Marymount University](#)) and Ana H Valentin Torres Dsc ([Marymount University](#)).

### ***NotPetya Wiper Malware Disrupted Business Operations.***

**Abstract.** The purpose of this poster is to describe how the NotPetya wiper malware attack occurred in 2017. The poster describes how the NotPetya wiper malware successfully impacted the user and Kernel modes exposing the security and privacy from major worldwide systems. The poster explains how the NotPetya wiper malware breaches the firewall of all security applications. The poster also identifies the elements of the intrusion detection systems and the mitigation strategies through different established frameworks.

Endre Szakal ([James Madison University](#)), Allison Tucker ([James Madison University](#)), Leah Wilczynski ([James Madison University](#)), Mia Pham ([James Madison University](#)), Mace Bentley ([James Madison University](#)), Zhojun Duan ([James Madison University](#)), Tobias Gerken ([James Madison University](#)), Dudley Bonsal ([James Madison University](#)) and Henry Way ([James Madison University](#)).

***Development of a lightning flash clustering algorithm to identify thunderstorm events within the Washington, DC metropolitan area.***

**Abstract.** With over seven million warm-season, cloud-to-ground lightning flashes in Washington, DC to analyze, producing a clustering algorithm capable of resolving individual thunderstorm events completes an important step in the analysis and visualization of lightning data. A thunderstorm event search algorithm was generated following the lightning clustering rules within the study; Identification and analysis of flash cells in thunderstorms (Tuomi and Larjavaara,2004). The algorithm was calibrated using specific parameters aimed at discerning a wide temporal and spatial range of individual thunderstorm events from lightning detected by the National Lightning Detection Network (NLDN). With lightning flashes processed in a time-order sequence, the algorithm assigned flashes to events based on whether the differences between their spatial and temporal attributes met certain thresholds. After being run on a Washington DC lightning flash dataset from 2006 to 2020, the algorithm identified 173,817 thunderstorm events. Because input parameters are variable and can cause the algorithm to produce different results, further calibration was necessary to ensure an adequate identification resolution was found for resolving thunderstorm events. Complications of thunderstorm identification from lightning flash data included overlapping thunderstorm tracks, short-lived thunderstorms, and stationary thunderstorms. The resulting thunderstorm event clusters are relevant for a multitude of geographic analyses and visualizations.

## Flash Talks

Nasreen Muhammad Arif ([Old Dominion University](#)).

### ***Data Science in Action: The Impact of Vaccination and Age on COVID-19 Cases and Deaths.***

**Abstract.** This scientific research is about correlating the COVID-19 cases and death data with peoples' vaccine status and age and finding the relation between the two datasets and finding informative insights from it. The data is extracted from the case surveillance of the United States (US) website. The primary data represents the US COVID-19 Cases and Deaths by State and COVID-19 Vaccinations in US County as a secondary dataset. COVID-19 has made a huge impact all over the world since 2019. Scientists have made some vaccines for their recovery and their behavior is different for each individual. COVID-19 vaccine status has a huge impact on the COVID-19 related cases and deaths. This study can benefit researchers who are working on COVID-19 analysis. This scientific research can help the scientific community and policy makers to analyze the impact of vaccines on the COVID-19 situation. This study may help prepare test and training data to train the model in order to predict COVID-19 cases, deaths, and recovery rate. The data analysis in this study was performed using Python in the Visual Studio Code environment, and pandas' visualization is built using matplotlib and seaborn libraries. Machine learning algorithms are used in Weka for data training.

Lexi Burroughs ([Christopher Newport University](#)).

### ***Teachings from Tech Executives.***

**Abstract.** What if you could hear the best advice top technology executives have to give from the comfort of your home, in just five minutes? My school's Society of Women Engineers chapter invited multiple highly-qualified executives in the technology industry to give presentations surrounding their careers, lessons learned, and more. In this flash talk, I will cover the biggest takeaways from each of these executives' presentations. These takeaways include the potential benefits of a non-linear career path, why mistakes are good (and the biggest mistake/accident that happened to them on the job), how to break into the industry if you don't have a background in computing, and more. Additionally, with four of the five executives being women, this flash talk will discuss their experiences being high-profile women in a male-dominated industry, resources for networking with other professional women, and how to find power in your voice. By shortening the teachings of executives into bite-sized pieces, anyone from any background can quickly glean insight on how to become successful in the technology industry and what resources there are within the technology space.

Megan Caulfield ([James Madison University](#)) and Jason Forsyth ([James Madison University](#)).

### ***Wearable Computing for Physical Rehabilitation.***

**Abstract.** According to the National Center for Biotechnology, adherence to at-home exercise in rehabilitation is a large issue, finding that as high as 50% of patients are not following their at-home plans. Physical therapy plays a significant part in the rehabilitation process; including an implementation of exercises to perform at home. Patients who fail to follow through with their at-home exercises potentially can face detrimental effects to their recovery. The proposed research seeks to explore wearable computing software to develop a device that enables physical therapy patients to receive feedback on their movements while at home and apply corrections as though a therapist was working hands-on with them.



Wearable technology is useful because it is directly on the body allowing us to collect user data without infrastructure being installed in an environment. These systems are challenging to develop as the device must fit the user comfortably, and be adjustable and durable. In our current work, we are exploring haptic feedback for simple activities. We have begun developing a glove with motors embedded inside that will vibrate guiding the movement of the patient from one position to another. We plan to conduct a user study where we will test the prototypes using haptic and visual feedback to determine what the preferred feedback method might be. The end goal of this research is to design a wearable device that can identify a patient's posture, interpret how to improve their motion, and provide feedback as to how the patient can improve.

Swaroop Dola ([University of Virginia](#)), Matthew B. Dwyer ([University of Virginia](#)) and Mary Lou Soffa ([University of Virginia](#)).

### ***Input Distribution Coverage: Measuring Feature Interaction Adequacy in Neural Network Testing.***

**Abstract.** Testing deep neural networks (DNNs) has garnered great interest in the recent years due to their use in many applications. Black-box test adequacy measures are useful for guiding the testing process in covering the input domain. However, the absence of input specifications makes it challenging to apply black-box test adequacy measures in DNN testing. The Input Distribution Coverage (IDC) framework addresses this challenge by using a variational autoencoder to learn a low dimensional latent representation of the input distribution, and then using that latent space as a coverage domain for testing. IDC applies combinatorial interaction testing on a partitioning of the latent space to measure test adequacy.

Empirical evaluation demonstrates that IDC is cost-effective, capable of detecting feature diversity in test inputs, and more sensitive than prior work to test inputs generated using different DNN test generation methods. The findings demonstrate that IDC overcomes several limitations of white-box DNN coverage approaches by avoiding infeasible inputs in testing and enabling the calculation of test adequacy metrics that capture the diversity present in the input space of DNNs.

Jinsun Lee ([Freeman High School](#)), You Jin Lee ([Brown University](#)) and Wookjin Choi ([Thomas Jefferson University](#)).

### ***Covid-19 Data Analysis.***

**Abstract.** Data Analytics is a broad field of study and a profession that, at its core, addresses the growing amount of technology and corresponding data produced from the multitude of data sources. While data potentially provides scientific and rational ways for people to make informed decisions, the usefulness of a massive amount of data fundamentally depends on analytics. Analytics is how one makes sense of the data and draws meaningful insights.

Since the COVID-19 outbreak, hospitals, universities, and government agencies have made COVID-19 datasets publicly available. While the COVID-19 datasets are the crucial sources for understanding the pandemic, they remain cryptic to most laypeople. In this work, we analyzed the COVID-19 datasets available on the Virginia Department of Health website. The goal of the project was twofold. First, we aimed to draw insights from visualizing the COVID-19 dataset. Second, we wanted to see how accessible the public COVID-19 datasets are to the students who are still learning data analytics techniques.



We accessed the datasets using Python on Google Colab. The data curation was a big part of the work. After importing the datasets onto the Pandas dataframes, we had to clean the dataset before conducting any analysis. After visualizing the cleaned data, we compared the graphs with known past events such as Delta and Omicron variation outbreaks, holiday seasons, and vaccination status of the general population.

Matthew Marini ([James Madison University](#)), Tia Stamp-Querry ([James Madison University](#)), Chris Ondoua ([James Madison University](#)), Sherinne Zhang ([James Madison University](#)), Jack Navarette ([James Madison University](#)) and Shraddha Joshi ([James Madison University](#)).

### ***Relationships with Internet of Things.***

**Abstract.** There are obstacles in today's world that prevent people from being able to see each other and experience moments of true connection that are only possible when they are face-to-face. This could be long distance, busy schedules, and even recently the COVID-19 pandemic. The team's goal is to use IoT technology to design a tangible device to allow for a more meaningful connection than video calls and texts provide.

The focus of our two-year capstone experience is to help siblings maintain and enhance connections when they are physically separated. More specifically, we are concentrating on the connection between college students with their siblings who are between their early teen years to adult years. We plan to design a product using the Internet of Things (IoT) technology to alleviate the pains of not being able to connect with siblings face-to-face.

IoT describes the network of physical objects that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet.

We have transitioned into establishing design requirements and starting the concept generation phase in our Spring 2022 semester. The team will discuss their rationale through the definition phase of the design process and how the stakeholder, problem, and scope were defined. After this, the ideation phase will be discussed along with the methods used for concept generation and selection. The final two semesters of the project will consist of fabricating and testing our design to meet our stakeholder's needs.

Takiva Richardson ([Virginia State University](#)), Joseph Shelton ([Virginia State University](#)), Yasmin Eady ([North Carolina A&T State University](#)), Kofi Kyei ([North Carolina A&T SU](#)) and Albert Esterline ([Dept. of Computer Science, North Carolina A&T State University](#)).

### ***WebID + Biometrics with Permuted Disposable Features.***

**Abstract.** For networked communications, cyber security and authentication are critical components. This work deals with the issue of security and authentication as it relates to the Semantic Web. The phrase "Semantic Web" alludes to the World Wide Web Consortium's idea of standards to make internet data machine-readable and reusable. People may use Semantic Web technologies to establish online data repositories, define vocabularies, and implement data-handling rules. WebID, for example, is a technique for managing profile data connected with people and services at self-defined locations. Although this gives users greater control over their data, networked connections are still vulnerable to replay assaults. While password-based authentication is the most frequent method, biometric-based authentication is becoming more popular as biometrics cannot be forgotten or stolen as readily as a password. Successful replay attacks on a biometrics-based authentication system, on the other hand, are particularly damaging since biometrics are more difficult to change than passwords. Prior work has been done to develop unique and accurate representations of one's biometric, such that if biometric data is collected, it will not be useful in attacking the system because the data is only used once. While the suggested approach showed potential, the number of distinct representations that could be constructed restricted it. In this paper, we suggest extending previous work to increase the number of unique and reliable data representations. This paper will provide an overview of the system that is currently under development, as well as additional WebID components.

Zahra Rizvi ([York High School](#)).

### ***Eye and Head Movement Tracking: An Innovative Real-time Detection of Driver Distraction using Machine Learning.***

**Abstract.** The use of cell phones while driving results in distracted driving. It is a growing public safety hazard. The motivation of this scientific research is to bring awareness and scientifically demonstrate the risks of distracted driving. It highlights the benefits of real-time tracking eye and head movement using a camera. This science experiment focuses on visual distraction during which drivers' eyes are off the road. In this research, a new computer program was written in Python that uses machine learning algorithms to track eyes and head movement using a camera. The program uses OpenCV to detect face, head and eyes. Also, TensorFlow is used as a backend for Keras which is used for classification modeling. The camera first captures the face of the motorist. Next, after detecting the face, the program creates a region of interest (ROI). The eyes are then detected from ROI. This is fed to the classifier. The classifier then determines if the eyes are on or off the road. It then records the status of eyes with a timestamp in a log file. The log file is used to get the output data that may be used for data analysis. The research is particularly relevant in the post coronavirus pandemic era, given that millions of people around the world have taken to heavily using the popular mobile apps, and are tempted to use them while driving. This study will help spread awareness about the serious and adverse consequences of distracted driving.

Donna Schaeffer ([Marymount University](#)).

### ***Design Thinking: A Framework For Teaching Cybersecurity by Design.***

**Abstract.** The purpose of this flash talk is to introduce design thinking as a framework for solving cybersecurity problems. For example, design thinking concepts can allow designers to seamlessly integrate cybersecurity controls into the user's environment. Another example is to include users early in the design process so that they buy into and feel invested in the importance of compliance with cybersecurity controls.

Students and faculty may become more aware of the design thinking technique, which has roots outside of the traditional computer science discipline, thus introducing a multidisciplinary approach. I will present the steps in this iterative process, and demonstrate one use case where design thinking is used in a cybersecurity program.

Johanna Schetelig ([Randolph-Macon College](#)) and Robin Givens ([Randolph-Macon College](#)).

### ***Exploring Machine Learning Algorithms for Emotion Recognition Through Optimized Input Training Datasets.***

**Abstract.** Automated facial emotion recognition has become increasingly relevant for human-computer interaction and relies on the efficient implementation of machine learning algorithms to correctly identify complex patterns in human faces. This work compares the combined performance of three algorithms, Support Vector Machines (SVC), Random Forest Classifiers, and k-Nearest Neighbor (k-NN) as an alternative to computationally expensive neural networks. To identify facial patterns associated with an emotion, facial landmarks were recognized, and their distances were computed. These were then used as input for the models. The performance of these models was evaluated based on efficiency and accuracy with varying training datasets. Based on an existing facial emotion recognition dataset (FER-2013), subsets were created to be used as training to overcome an issue of misclassification due to overlapping expression features in multiple different emotions. Each model was then trained on a different dataset with unique facial features associated with a given emotion.

Combining the predictions of the three models it was possible to increase the overall accuracy while maintaining a low training time. Despite having significant difficulties recognizing subtle details in facial expressions compared to the neural network, our combined models managed to accurately recognize strong expressions associated with each emotion.

## Birds of A Feather

Ariel Arison ([University of Mary Washington](#)).

### ***Leading Student Clubs.***

**Abstract.** Support throughout college is invaluable especially for minorities within a field. Collegial student clubs promoting diversity and inclusion can provide support for members through organizing programming and events. However, it is challenging to develop ideas for events and decide what would be most beneficial. This discussion will focus on exchanging ideas between executive officers and club members from different institutions.

Amy Vaillancourt ([University of Maryland](#)).

### ***Meet Me at the Intersection of Gender and Disability.***

**Abstract.** This Birds-of-a-feather session is an invitation for women/nonbinary folks with disabilities and their accomplices to consider the question; how does the computing education and tech industry identify and support the needs of women and nonbinary folks with disabilities? We will consider definitions of the relevant terms, share experiences, and identify questions and assumptions we may have as community members and accomplices. Additionally, we will explore the consumer needs of the community as defined by the community and where tech education and the tech industry meets and does not meet those needs.

The intended audience is students, faculty, and industry professionals who are interested in how to build a more inclusive community for women and nonbinary individuals who identify as disabled.

Conversation starting questions include:

- Where does the tech community - both educators and industry - stand in terms of inclusion for disabled women in tech?
- What rights and responsibilities does a women/nonbinary and disabled community member have within the tech industry?
- How can educators and administrators best help prepare students from this community for work in the tech industry?
- What responsibilities does the larger tech community have to serve workers in this community?
- How does the tech industry serve consumers at the intersection of women/nonbinary folks and disability? What else could tech do to meet their needs?

The moderator is Amy Vaillancourt, Break Through Tech DC Community Outreach Coordinator, Academic Advisor and a member of the President's Commission on Disability Issues at the University of Maryland.

## Technical Workshops

Aisha Ali-Gombe ([Towson University](#)).

### ***The Basics of Mobile Malware Analysis.***

**Abstract.** Malware are programs that infect target systems and perform unwanted or illicit activities. In recent years, the Android platform has seen the proliferation of malicious software. Related literature has characterized Android malware based on different payload types such as information stealing, financial charges, privilege escalation, etc. This workshop will introduce participants to the basics of Android malware analysis and functionality detection using simple open-source and free static and automated analysis tools such as Androguard and Flowdroid.

After attending this workshop, participants will gain a basic understanding of Android malware characterization. They will learn how to statically dissect Android malware to identify potential red flags such as excessive permissions, sensitive APIs, native code, and data exfiltration.

We will provide access to a Virtual Machine disk configured with all required setup and analysis tools to all participants, which can be downloaded and run locally using any of the following virtualization software - VMware Player, Fusion, Workstation, Virtual Box. No prior cybersecurity knowledge is needed. A basic understanding of Java programming is essential.

## Technical Talks

Shubhi Asthana ([IBM Research](#)), Pawan Chowdhary ([IBM Research](#)) and Taiga Nakamura ([IBM Research](#)).

### ***Transforming the Order Management process.***

**Abstract.** The transactions of goods and services between large businesses are often driven by contracts and orders. Based on usage of services, invoices are billed to the customer who settle them. Enterprise service provider process thousands of orders every year and it takes considerable manual effort to process them. The invoice's billed data is not always maintained in the same system as the orders, and it takes effort to match them. Sometimes the invoices may get into a dispute due to over exhaustion of allocated funds or may be billed to an expired order. Hence managing the billing of order is a huge undertaking along with increased cost.

To address these challenges, we developed an order management workflow system, that transforms the overall orders process to increase their renewals and decrease disputes. It comprises of two components – an automated order-invoice model and a risk analytics model that evaluates the contracts and orders against the invoices billed to the customer. The output is the ranking of orders based on customer portfolio, risk level as well as market trends in usage of services. We illustrate our approach with some promising results on data of one of the world's largest IT service providers.

Daniel Broder ([Marymount University](#)).

### ***Leveraging AWS For Enterprise Data Using S3, Glue, Athena/Redshift and QuickSight.***

**Abstract.** In this technical talk, I will demonstrate how you can set up an enterprise data operation for your organization using AWS tools, including S3, Glue, Athena and QuickSight. We will go through the entire data lifecycle process, from inception (object appears in the S3 bucket) to cleaning (with Glue), processing (with Redshift or Athena) and analysis with QuickSight. This talk will be an entire technical demonstration with room for questions at the end.

Cheryl Howard ([IBM](#)).

### ***MLOps: Bringing AI and Machine Learning out of the Sandbox.***

**Abstract.** Machine Learning and Artificial Intelligence (ML/AI) technology has found its way out of the laboratory and into the mainstream across a vast range of applications. Although most enterprises understand the huge potential of ML/AI to optimize resources, increase customer loyalty and streamline business processes, many of these projects never emerge from the "sandbox" into production usage. MLOps is a set of best practices for building, deploying and maintaining machine learning models, and can be considered a specialization of the DevOps methodologies that have become standard in software development. This talk will delve into the MLOps pipeline and address pitfalls and best practices at every stage from use-case definition through model building, deployment, and monitoring models for fairness, explainability, and sustained performance.

## Sponsors / Career Fair

### ***Virginia Tech Graduate School***

The Virginia Tech Innovation Campus will be both a place and a culture that unlocks the power of diverse people and ideas to solve the world's most pressing problems through technology. We have set the goal to be the most diverse graduate tech program in the country. Masters degrees in Computer Science and Computer Engineering will include project based learning and strong industry and government partnerships. Accelerated Bachelor's to Masters (4+1) programs are available to fast track the degree. Apply at [vt.edu/masters](https://vt.edu/masters) and send questions to [techtalent@vt.edu](mailto:techtalent@vt.edu).

