

The 21st KOCSEA Technical Symposium 2021

November 6-7 2021

The University of Nevada, Las Vegas, NV

The Future of Computing



KOCSEA

Korean Computer Scientists and Engineers Association in America
재/미/한/인/정/보/과/학/자/협/회

The 21st KOCSEA Technical Symposium

November 6 - 7, 2021
Las Vegas, NV

Table of Contents

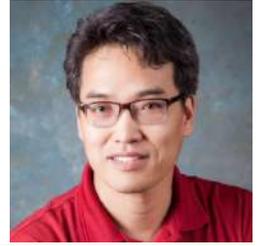
About KOCSEA	3
Message from the KOCSEA President	4
Message from the General Chair	5
Symposium Organizers	6
Technical Program	7
Keynotes	9
Abstracts Talks	12
Sponsors	21

About KOCSEA

The **K**Orean **C**omputer **S**cientists and **E**ngineers **A**ssociation in America (**KOCSEA**), founded in 1983, is a non-profit organization of Korean American computer scientists and engineers in the United States. The purpose of KOCSEA is to promote communication, information exchange and cooperation among its members and to provide opportunities for them to make contributions to computer science and related fields in Korea and the U.S.

KOCSEA consists of over 320 members at various stages in their career - from undergraduate and graduate students, to young professionals and assistant professors, to senior executives and professors. KOCSEA has organized annual workshops in Korea in cooperation with the Korean Information Science Society (KISS) from 1990 to 1995 in various areas of information technologies, including software engineering, multimedia technology, end user computing technology, and network technology. Recently, KOCSEA members have been actively participating in various technical activities promoting collaboration between US and Korea such as organizing symposiums at UKC (US Korea Conference on Science, Technology, and Entrepreneurship) of KSEA (Korea-American Scientists and Engineers Association), serving on the project proposal review panels of KOSEF (Korea Science and Engineering Foundation), and producing research trend study reports sponsored by KOSEF and KSEA in the area of computer and communications.

Message from the KOCSEA President



I am honored to welcome you to the 21st KOCSEA Technical Symposium. This annual symposium is our flagship event, gathering many Korean researchers and professionals in various computer and information technology fields. This symposium will provide the participants with an excellent opportunity to meet and network with fellow Korean and Korean-American computer scientists and engineers.

This year we have invited exceptional keynote speakers leading scholars in artificial intelligence, self-driving cars, data security, multimedia, and computer science education.

The technical sessions cover various vital areas for the future of intelligent computing. This year especially, we emphasize a professional networking opportunity. We are sure that these times will be valuable for everyone and hope that they will be fruitful.

We would like to thank our sponsors. The 21st KOCSEA Symposium and scholarships for Korean students studying computer science would not be possible without their generous support.

I also would like to express my sincere gratitude to the organizing committee members and volunteers who have devoted their time and energy to this fantastic annual event.

I hope you enjoy the symposium and make many meaningful connections for your research, teaching, and career in your fields. We look forward to meeting you at the symposium.

Sincerely,

Jaerock Kwon, Ph.D.
President, KOCSEA 2021

Message from the General Chair



I am honored to invite you to the 21st KOCSEA (Korean Computer Scientists and Engineers Association in America) Technical Symposium. This year, we will meet at Las Vegas, Nevada, on November 6th and 7th, 2021.

KOCSEA, founded in 1983, is a non-profit organization of Korean and Korean American computer scientists and engineers in North America. The goal of KOCSEA is to “promote communication, information exchanges, and cooperation among its members and to provide opportunities for them to make contributions to computer-related fields in the U.S. and Korea.” KOCSEA has several hundreds of members and patrons in academia and industry. They are the leading researchers and practitioners in the Computer Science and Engineering, Information Technology, and Data Science fields in the US.

The annual KOCSEA Technical Symposium has been very successful in the past years by providing opportunities to the participants in sharing their knowledge and information and promoting excellent networking among the Korean computer science community in the US.

The 21st Tech Symposium is a one-and-a-half-day event and consists of several keynote speeches by outstanding researchers and professors, many regular talks, and posters in various fields in computing fields. The symposium is a single-track conference to provide speakers excellent opportunities to promote their work in front of nearly one hundred participants from the US and Korea.

On behalf of the organizing committee members, I cordially invite you to the 21st KOCSEA Technical Symposium. Please feel free to bring your colleagues and students to the event. We look forward to meeting you at the symposium.

Sincerely,

Mingon Kang, Ph.D.
General Chair, KOCSEA Technical Symposium 2021
Vice President, KOCSEA 2021

Symposium Organizers

General Chair	Prof. Mingon Kang <i>University of Nevada, Las Vegas</i>
Program Chair	Prof. Kyu Hyung Lee <i>University of Georgia</i>
Sponsor Chair	Dr. Karl Kwon MITRE Prof. Yongwhi Kwon <i>University of Virginia</i>
Finance Chairs	Dr. Chul Sung IBM Research
Communication Co-Chairs	Prof. Donghoon Kim <i>Arkansas State University</i>
Scholarship & Membership Chairs	Prof. Sungchul Lee <i>Sun Moon University</i>
Registration Chair	Prof. Junggab Son <i>Kennesaw State University</i>
Publication Chair	Prof. Daeyoung Kim <i>Montclair State University</i>
Poster Chair	Prof. Doowon Kim <i>University of Tennessee, Knoxville</i>
Advisors	Prof. Jeho Park <i>Claremont McKenna College</i> Dr. Han Hee Song Prof. Seon Ho Kim <i>University of Southern California</i> Prof. Jeongkyu Lee <i>University of Bridgeport</i> Prof. Eunjee Song <i>Baylor University</i>

21st KOCSEA Technical Symposium Program

2021 KOCSEA Program: Day 1, November 6, 2021 (Saturday)

Session	Time	Speaker (Affiliation): Talk Title
Registration & Breakfast	08:00 - 08:40	Registration & Breakfast
Opening Remarks	08:40 - 08:50	Dr. Rama Venkat - UNLV Engineering Dean
Session 1: Cyber Security Meets Machine Learning Session Chair: TBA	09:00 - 10:00	- Hyungjoon (Kevin) Koo (Sungkyunkwan University), "Executable binary code representation with deep learning" ® - Haehyun Cho (Soongsil University), "Fuzzing the Binding Code of JavaScript Engines Using Semantically Correct Test Cases" ® - Chunghwan Kim (University of Texas at Dallas), "Securing Autonomous Vehicles through Software Testing and Analysis" ®
Break		Short Break
Keynote 1		Dr. Jung-woo Ha (NAVER AI Lab & CLOVA Research), "HyperCLOVA: Era of Hyperscale AI and research strategy of NAVER CLOVA" ®
Session 2: Machine Learning and Its Applications Session Chair: TBA		- Kangwook Lee (University of Wisconsin-Madison), "Information Theory and Coding for Trustworthy and Scalable Machine Learning" - Jeehee Lee (UNLV), "Data-driven Construction Dynamics" - Seon Ho Kim (University of Southern California), "Exploring the Spatial-Visual Locality of Geo-tagged Urban Street Images"
Lunch		Lunch & Poster Installation
Poster Session	01:30 - 03:30	23 posters will be presented
Short Break	03:30 - 03:50	Break
Keynote 2	03:50 - 04:30	Dr. Su-In Lee (University of Washington), "Explainable Artificial Intelligence for Biology and Health" ®
Session 3: Smart Communication and Security Session Chair: TBA	04:30 - 05:20	- Dam Hee Kim (University of Arizona), "Computer Science Meets Media and Communication: Social Media "Big Data" Analysis of Parasite" - Junghwan Rhee (University of Central Oklahoma), "Find My Sloths: Automated Comparative Analysis of How Real Enterprise Computers Keep Up with the Software Update Races"
Sponsor Talks Session Chair: TBA	5:20 - 05:40	Seok-Ho Kim (KUSCO), "Korea's Invitation Programs for Outstanding Overseas Researchers"
Banquet	05:40 - 09:00	Banquet with Scholarship Award Ceremony

® indicates remote presentation.

2021 KOCSEA Program: Day 2, November 7, 2021 (Sunday)

Session	Time	Speaker (Affiliation): Talk Title
Registration & Breakfast	09:00 - 10:30	Registration and Breakfast
Keynote 3	10:30 - 11:10	Dr. Shashi Nambisan (UNLV), TBD
Session 5: Data Analysis and Measurement Session Chair: TBA	11:10 - 12:10	- Yen-soon Kim (UNLV), TBD - Mingon Kang (UNLV), "Interpretable deep learning enhances prediction of enzyme commission numbers" - Chul-Ho Lee (Texas State University), "Transient Dynamics of Epidemic Spreading and Its Mitigation on Large Networks" ®
Poster Session	12:10 - 1:30	23 Posters will be presented *
Closing Remarks and lunch	01:30 - 01:40	

® indicates remote presentation.

Poster Presentation

- Joon-Ho Choi, "Smart Buildings as a Function of Human-Building Integration"
- Changhun Jo, "Discrete-Valued Latent Preference Matrix Estimation with Graph Side Information"
- Tae Oh, "An Exploratory Study to Examine the Reasons for the Low Adoption Rate of Smart Canes"
- Alireza Khamesipour, "Classification of microarray data based on pairwise Biomarker discovery"
- Sam Chung, "Cloud-Based Software Reverse Engineering"
- Sejun Song, "Apple In My Eyes (AIME): Liveness Detection for MobileSecurity Using Corneal Specular Reflections"
- Ahyoung Lee, "Experimental Network Traffic Load Testing in Cloud Environments"
- Gichan Lee, "Interpreting Machine Learning based Code Smell Severity Prediction with Explainable AI"
- Jemin Ahn, "Rapid Web-Page Loading with Preload Scanner"
- Ilho Noh, "Joint learning for biomedical NER and entity normalization"
- Brian Maeng, "Coordinating an Emission-dependent Supply Chain under Cap-and-trade Regulation"
- Jongyong Park, "Low cost Motion Capture System"
- Sang Kim, "Low-Latency Detection of False Data in Short Packets"
- Sooyong Jang, "Improving Classifier Confidence using Lossy Label-Invariant Transformations"
- Jungseok Hong, "Semantically-Aware Strategies for Stereo-Visual Robotic Obstacle Avoidance"
- Sai Kosaraju, "HipoMap: Graphical Representation Map for WholeSlide-based Histopathological Image Analysis"
- Chijung Jung, "Swarmbug: Debugging Configuration Bugs in Swarm Robotics"
- Sian Lee, "Effects of Associative Inference on People's Susceptibility to Fake News: Experimental and Empirical Evidences"
- Sian Lee, "Optimal auction order with two different items in sequential Vickrey auctions with reserve price"
- Kijoon Bang, "Bayesian reaction optimization in solution based thermal reaction toward single crystalline perovskite semiconductor"
- Bumsu Kim, "Curvature-Aware Derivative Free Optimization"
- Junho Hong, "Cybersecurity of EV charging station"
- Euseong Ko, "Sex-specific Risk Score Model with Asthma Data"

Keynotes

1. **Dr. Jung-Woo Ha**

Head of NAVER CLOVA AI Lab,

HyperCLOVA: Era of Hyperscale AI and research strategy of NAVER CLOVA



Dr. Jung-Woo Ha is the head of NAVER CLOVA AI Lab, responsible for AI research in NAVER. He works as the co-directors of two AI collaboration research centers: SNU-NAVER Hyperscale AI Center and KAIST-NAVER Hypercreative AI Center. He also coordinates company-level AI R&D collaboration as the chair of NAVER AI Tech Committee. In addition, he coordinates AI research collaboration between international research groups, such as NAVER LABS Europe, NYU, CMU, Univ. of Tuebingen, HUST, PTIT, and Wattpad for NAVER Global AI R&D Belt. Dr. Ha got his BS and PhD degrees from Seoul National University in the major of machine learning. During the last three years, he has published over 20 papers at top-tier AI conferences, such as NeurIPS, ICRL, CVPR, ACL, ICASSP, AAAI, and so on. Also, He has served as the reviewer or program committee member of those conferences. His research interests include deep learning-based computer vision, natural language processing, multimodal representation, audio signal modeling, and their real-world applications. In particular, he focuses on hyperscale representation learning approaches and their extension to various domains and applications.

Abstract: Recent AI research trends can be explained by the keyword “hyperscale” in model, data, and infrastructure. What made this possible is the explosive growth of multimodal data with advances in the Internet, mobile, and various user device technologies, and hardware and infrastructure technologies. In addition, the advancement of self-supervised learning has contributed mainly to learning explosive data without human annotations. In particular, since OpenAI released GPT-3, it has shocked many AI researchers around the world. In particular, GPT-3 with 175 billion parameters is said to be changing the existing learning paradigm by showing comparable performance with SOTA methods through in-context few-shot learning only with prompts without additional fine-tuning learning. But, unfortunately, most of the corpus for training GPT-3 is English data, and the proportion of Korean is only 0.016%, making it impossible to use it for Korean language users. Considering the outstanding performance of GPT-3, it was necessary to build a Korean-centered language model with more than tens of billions of parameters to provide innovative technologies and AI tools to domestic users. As a technology platform company, Naver developed HyperCLOVA, a hyperscale Korean-centric language model, and released the results in May 2021 through large-scale infrastructure investment, Korean-centered data construction, and research from September 2020. This lecture introduces the latest hyperscale AI technology research trend and presents various cases of efforts and future directions for competition in the era of hyperscale AI, focusing on NAVER CLOVA's HyperCLOVA.

2. Dr. Su-In Lee

Paul G. Allen Professor
the Paul G. Allen School of Computer Science & Engineering
University of Washington



Explainable Artificial Intelligence for Biology and Health

Abstract:

Her research has conceptually and fundamentally advanced how AI can be integrated with biomedicine by addressing novel, forward-looking, and stimulating questions, enabled by AI's potential. For example, although the primary focus of AI applications in the field of medicine had been on accurately predicting a patient's phenotype or outcome, she focused on the question of why. This line of work has led to highly cited seminal publications in the field of foundational AI, clinical medicine, and computational molecular biology. Her research aims to push the boundaries of both foundational AI and molecular biomedicine, to address new questions and make novel discoveries from high-throughput molecular data or patient's medical record data.

Short Bio:

Dr. Su-In Lee is a Paul G. Allen Professor in the Paul G. Allen School of Computer Science & Engineering at the University of Washington. She completed her PhD in 2009 at Stanford University with Prof. Daphne Koller in the Stanford Artificial Intelligence Laboratory. Before joining the UW in 2010, Lee was a Visiting Assistant Professor in the Computational Biology Department at Carnegie Mellon University School of Computer Science. She has received the National Science Foundation CAREER Award and been named an American Cancer Society Research Scholar. She has received generous grants from the National Institutes of Health, the National Science Foundation, and the American Cancer Society.

3. Dr. Shashi Nambisan

Director, Transportation Research Center and Professor of Civil Engineering,
University of Nevada, Las Vegas, USA

**Data and Analytics for Decision Support: Some Considerations in the Transportation Domain****Abstract:**

The presentation will address some considerations related to leveraging data for decision support in the transportation domain. It will highlight needs, challenges, and opportunities particularly with regard to technology considerations and operational ecosystems that include Connected and Automated Vehicles and Infrastructure Systems (CAVIS) and Automated, Connected, Electric, and Shared Systems (ACES) with a focus on safety. These include data needs, considerations, and concerns from the perspectives of infrastructure system owners, operators, managers, and users. Data elements of interest include extent and conditions of the infrastructure, user and vehicle characteristics, user-vehicle-system interactions, and safety. Examples include temporal and spatial resolution, accuracy, timeliness, relevance, risks, and threats.

Short Bio:

Dr. Shashi Nambisan, P.E., is the Director of the Transportation Research Center and a Professor of Civil Engineering at the University of Nevada, Las Vegas. He has more than 32 years of experience including roles in Nevada, Iowa, Tennessee, and Alabama. As Principal Investigator (PI) or co-PI, he has led efforts on more than 170 projects to support decision making that span the transportation system life cycle. He has addressed a broad range of transportation topics with safety and data-enabled decision support being a prevailing theme. His students have received numerous scholarships and awards at local, regional and national levels, and many of them serve in professional leadership roles. Among the awards and honors he has received is a proclamation by the Governor of the State of Nevada designating January 31, 2007 as the "Professor Shashi Nambisan Day" in recognition of his leadership role in and contributions to transportation in Nevada.

Abstracts | Talks

Shown in the order of presentation

1. **Hyungjoon (Kevin) Koo (Sungkyunkwan University), "Executable binary code representation with deep learning"**

Software is everywhere. The distribution format of software is an executable binary. After a complex compilation process, it is inevitable to lose varying information available in the original program source such as variable names, object types, function arguments, and so forth. Recent advances with deep neural networks (particularly RNN or NLP) shed light on the possibility of recovering such disappeared information. Predicting the information that has been lost in executable binaries can be applied to lots of areas including programmer authorship, binary similarity, code clone detection, malware family classification, etc. In this talk, we discuss what should be considered to represent executable binary codes for feeding a neural network model. Besides, we explore a few applications with such binary code representation.

2. **Haehyun Cho (Soongsil University), "Fuzzing the Binding Code of JavaScript Engines Using Semantically Correct Test Cases"**

JavaScript runtime systems include some specialized programming interfaces, called binding layers. Binding layers translate data representations between JavaScript and unsafe low-level languages, such as C and C++, by converting data between different types. Due to the wide adoption of JavaScript (and JavaScript engines) in the entire computing ecosystem, discovering bugs in JavaScript binding layers is critical. Nonetheless, existing JavaScript fuzzers cannot adequately fuzz binding layers due to two major challenges: Generating syntactically and semantically correct test cases and reducing the size of the input space for fuzzing. In this talk, I'm going to introduce Favocado, a novel fuzzing approach that focuses on fuzzing binding layers of JavaScript runtime systems. Favocado can generate syntactically and semantically correct JavaScript test cases through the use of extracted semantic information and careful maintenance of execution states. This way, test cases that Favocado generates do not raise unintended runtime exceptions, which substantially increases the chance of triggering binding code. Additionally, exploiting a unique feature (relative isolation) of binding layers, Favocado significantly reduces the size of the fuzzing input space by splitting DOM objects into equivalence classes and focusing fuzzing within each equivalence class. We demonstrate the effectiveness of Favocado in our experiments and show that Favocado outperforms a state-of-the-art DOM fuzzer. Finally, during the evaluation, we found 61 previously unknown bugs in four JavaScript runtime systems (Adobe Acrobat Reader, Foxit PDF Reader, Chromium, and WebKit). 33 of these bugs were security vulnerabilities

3. **Chunghwan Kim (University of Texas at Dallas), "Securing Autonomous Vehicles through Software Testing and Analysis"**

Autonomous vehicles (AVs) such as drones and self-driving cars are rapidly emerging types of cyber-physical systems. With their increasing adoption, AVs are facing threats of cyber and cyber-physical attacks that exploit their vulnerabilities. Although many AVs are critical to human safety and the environment, it is difficult to secure them against such attacks due to new challenges associated with AV software, such as constantly changing input data from various sensors and physical operations to maneuver the vehicle in real time. In this

talk, I will present an overview of my research that discovers and analyzes AV software vulnerabilities. Such vulnerabilities can be exploited by attackers to maliciously control the AV and cause a severe vehicle accident (e.g., a crash). I will first discuss a new type of such vulnerabilities in drone control software and how we have found these vulnerabilities through automated software testing. I will then introduce how we analyze the control software and vehicle logs to identify the root cause of such attacks after an accident happens. Finally, I will conclude this talk by discussing my ongoing work on finding and analyzing software vulnerabilities in self-driving cars.

4. Kangwook Lee (University of Wisconsin-Madison), "Information Theory and Coding for Trustworthy and Scalable Machine Learning"

In this talk, I will present the role of information theory and coding for enabling large-scale trustworthy ML systems. In the first part, I will talk about the role of information theory in developing a holistic framework for fair and robust machine learning. In the second part, the role of coding in building large-scale machine learning systems will be presented. I will talk about the recent developments in coded computation, a principled interface between distributed systems and coding theory.

5. Jeehee Lee (UNLV), "Data-driven Construction Dynamics"

In an era where Big Data is transforming information sciences, data analytics contribute to major innovation by providing exciting opportunities for substantial breakthroughs and more solid decisions in the construction industry through its various phases. Given the nature of the human experience-oriented construction and civil engineering industry, useful data comes from everywhere – from projects to the buildings themselves. However, most of the data in the construction industry is unstructured format (e.g., free-form text, printed documents, analog sensor data, etc.) and it requires data-driven and data-centric approaches to obtain meaningful knowledge from the data. This talk centers on clarifying construction projects' uncertainties using multimodal text analytics as a driving force of change when connecting the construction project to every process, phase and stakeholders.

6. Seon Ho Kim (University of Southern California), "Exploring the Spatial-Visual Locality of Geo-tagged Urban Street Images"

Urban street images have a unique property as they capture visual scenes that are distinctive to their geographical regions. Such images are similar to their neighboring ones while dissimilar to faraway images. We refer to this characteristic of images as the spatial visual locality or the spatial locality of similar visual features. This study focuses on geo-tagged urban street images and hypothesizes that those images demonstrate a local similarity in a certain region but a dissimilarity across different regions, and provides different analysis methods to validate the hypothesis. The paper also evaluates the correctness of the hypothesis using three real geo-tagged street images collected from the Google Street View. Our experimental results demonstrate a high locality of similar visual features among urban street images.

7. Dam Hee Kim (University of Arizona), "Computer Science Meets Media and Communication: Social Media "Big Data" Analysis of Parasite"

Dam Hee Kim will share insights into how scholars in the fields of Computer Science and Media & Communication can conduct fruitful collaborative research. For this, she will draw from her recent collaborative project analyzing English vs. Korean tweets about the Oscar-winning Korean film, Parasite. First of all, with the unprecedented popularity of Parasite, Squid Game, and BTS, there is a growing need for research on what makes Korean entertainment content uniquely successful in the global market. This question can be answered by scholars with data analytical skills along with intercultural expertise. Second, we can conduct better quality research in an effective way when it is guided by relevant social theories involving media and communication. Lastly, computerized analyses of "big data" can be complemented and improved upon by in-depth, nuanced analyses by humans to draw valid conclusions.

8. Junghwan Rhee (University of Central Oklahoma), "Find My Sloths: Automated Comparative Analysis of How Real Enterprise Computers Keep Up with the Software Update Races"

A software update is a critical but complicated part of software security. Its delay poses risks due to vulnerabilities and defects of software. Despite the high demand to shorten the update lag and keep the

software up-to-date, software updates involve factors such as human behavior, program configurations, and system policies, adding variety in the updates of software. Investigating these factors in a real environment poses significant challenges such as the knowledge of software release schedules from the software vendors and the deployment times of programs in each user's machine. Obtaining software release plans requires information from vendors which is not typically available to public. On the users' side, tracking each software's exact update installation is required to determine the accurate update delay. Currently, a scalable and systematic approach is missing to analyze these two sides' views of a comprehensive set of software. We performed a long term system-wide study of update behavior for all software running in an enterprise by translating the operating system logs from enterprise machines into graphs of binary executable updates showing their complex, and individualized updates in the environment. Our comparative analysis locates risky machines and software with belated or dormant updates falling behind others within an enterprise without relying on any third-party or domain knowledge, providing new observations and opportunities for improvement of software updates. Our evaluation analyzes real data from 113,675 unique programs used by 774 computers over 3 years.

9. **Fatih Demir (Northern Illinois University), "Adopting Emergent Technologies for Teaching and Learning and User-Centered Design"**

Many recent studies have suggested that the use of technology in the classroom can help ESL students to acquire language skills at a faster rate (Yang & Walker, 2015). Extensive research has been done to show that Mobile Assisted Language Learning (MALL) second language (L2) learning is considered favorably by students (Darsih & Asikin, 2020), can be used effectively in authentic environments (Shadiev et al., 2017) and increases proficiency in language learning by multiple indicators (Nishida et al., 2012).

We sampled 20 ESL students ranging from ages 9-12. The length of time each student had spent in ESL courses ranged from less than one year to more than four years. The study involved students using a voice assistant system to answer a set of questions, conducting a usability test, and interviewing each student. Software was used to time the students and to record the answers to the question set. The data was analyzed using spreadsheet software. To measure effectiveness, the percentage of correct answers was calculated. To measure efficiency, the amount of time it took to complete the questions was calculated. To measure subjective satisfaction the students were administered the system usability scale (SUS).

Overall, the effectiveness, efficiency and user satisfaction of voice assistant systems make them a useful supplementary education tool for young ESL students as they are acquiring and using English. Being equipped with such a tool may help alleviate some of the anxiety that comes with the communication challenges of learning a second language.

10 **Yen-Soon Kim (UNLV), "Robot Application in Hospitality and the Needs: Developer and User Perspectives"**

The COVID pandemic has leaded traditional hospitality experience into contactless service. Hospitality firms have explored opportunities to incorporate technologies, such as robotics, into their production or service. Current applied robot technologies and the hospitality market needs as user perspectives would be introduced and how computer science and mechanical engineers can develop new robot product and the application for hospitality professionals and their customers. Although current may academic studies state that service or collaborative robots may be used for creating "Cool" experiences in hospitality some activities are not acceptable for robotization and it could lead to a potential loss of human contact or service and customers' dissatisfaction with the dining experience. This presentation will provide a real Bobarobo story, the first milk tea barista robot in the US market, including what things hospitality professionals are looking for as a developer and as well as customer perspectives. Regardless of advances in social robotics, still the application of current AI robots in the hospitality industry (particularly in F&B service) has been extremely limited.

11 **Chul-Ho Lee (Texas State University), "Transient Dynamics of Epidemic Spreading and Its Mitigation on Large Networks"**

This talk will present our recent work on the transient dynamics of the susceptible-infected (SI) epidemic spreading process on a large network. The SI model is naturally a better fit for modeling malware propagation in early times when patches/vaccines are unavailable or over a wider range of timescales when massive patching is practically infeasible. However, it has been largely overlooked in the literature. Its

analysis is also non-trivial, as its important dynamics are all transient, and the usual stability/steady-state analysis no longer applies. In this talk, I will explain why the transient dynamics are just as important as the steady-state behaviors and how they can be better estimated via our new theoretical framework. I will also discuss its application on the development of vaccination policies to mitigate the epidemic spreading to the extent possible.

12 Mingon Kang (University of Nevada, Las Vegas), "Interpretable deep learning enhances prediction of enzyme commission numbers"

The Enzyme Commission (EC) numbering system is widely used to classify enzyme functions. As of March 2021, 564,277 (manually curated) and 207,800,733 (computationally annotated) numbers of protein sequences have been identified as their EC numbers in the Uniprot database. However, there are still a large number of protein sequences, which are unknown their EC numbers in this database⁶. Furthermore, along with the advance of high-throughput genome sequencing technologies and environmental changes, an increasing number of new species of organisms (e.g., microorganisms, fungi, and plant) has been reported, where there are substantial unknown proteins. We developed an interpretable hierarchical deep learning model, named ECPICK, that can accurately classify about 5000 EC numbers. More importantly, ECPICK can identify potential active sites that may be important regions to classify the EC numbers. The identification of active sites is promising for biological discovery.

13 Joon-Ho Choi (University of Southern California), "Smart Buildings as a Function of Human-Building Integration"

A series of human subject experiments were performed in an environmental chamber at USC to collect different types of data, such as human body indexes, including skin temperature, electrode activity, and heart rate variability. The last two represent the occupants' reflections of outside stimuli and their stress levels, respectively indexes. Environmental parameters including indoor temperature were also defined. The study used machine learning to construct an algorithm set, and used it to build an HVAC auto-control process based on real-time sensor data.

14 Changhun Jo (UW Madison), "Discrete-Valued Latent Preference Matrix Estimation with Graph Side Information"

Incorporating graph side information into recommender systems has been widely used to better predict ratings, but relatively few works have focused on theoretical guarantees. Ahn et al. (2018) firstly characterized the optimal sample complexity in the presence of graph side information, but the results are limited due to strict, unrealistic assumptions made on the unknown latent preference matrix and the structure of user clusters. In this work, we propose a new model in which 1) the unknown latent preference matrix can have any discrete values, and 2) users can be clustered into multiple clusters, thereby relaxing the assumptions made in prior work. Under this new model, we fully characterize the optimal sample complexity and develop a computationally-efficient algorithm that matches the optimal sample complexity. Our algorithm is robust to model errors and outperforms the existing algorithms in terms of prediction performance on both synthetic and real data.

15 Tae Oh (Rochester Institute of Technology), "An Exploratory Study to Examine the Reasons for the Low Adoption Rate of Smart Canes"

"People who are blind and visually impaired (BVI) primarily rely on white canes for everyday mobility. With the advancement of technology, there has been an increased interest in researching more efficient modes of mobility since the white cane has several limitations. Smart canes were then developed to address the inadequacy of the white canes and to inform users of their surroundings more efficiently. While the smart cane's advantages are evident, the adoption rate of these devices is minimal due to several reasons. This paper explores different possibilities that may be contributing to the low adoption rate and aims to provide recommendations to increase smart cane awareness. We conducted an

exploratory user study by collecting survey data and interviewing participants and stakeholders in the BVI community. We analyzed both qualitative and quantitative data to identify potential reasons. Recommendations were made to address awareness and adoption concerns."

16 Alireza Khamesipour (City University of Seattle), "Classification of microarray data based on pairwise Biomarker discovery"

Simplicity, accuracy, and interpretability of decision rule systems for classification and class prediction of gene expression profiles are three quality metrics for determining the practicality of a classifier in the field. In our study, we introduce the AUC-based Top Scoring Pair (TSP) classifier, in which the calculated probabilities are based on the relative rankings of the two marker genes across all subjects as opposed to for each individual subject as proposed in the original TSP. Through extensive simulation results and case studies involving classification in ovarian, leukemia, colon, and breast and prostate cancers and diffuse large b-cell lymphoma, we show the superiority of the proposed approach in terms of improving classification accuracy, avoiding overfitting and being less prone to selecting non-informative pivot genes. The proposed AUCTSP is a simple yet reliable and robust rank-based classifier for gene expression classification. While the AUCTSP works by the same principle as TSP, its ability to determine the top scoring gene pair based on the relative rankings of two marker genes across all subjects as opposed to each individual subject significantly enhances performance in classification accuracy. The practicality of the selected gene pairs has been studied and justified using previous relevant research which emphasizes on the importance of AUCTSP features such as avoiding the selection of non-informative (pivot) genes as members of the TSP.

17 Sam Chung (City University of Seattle), "Cloud-Based Software Reverse Engineering"

Traditionally, companies have hosted applications locally or rented servers from other service providers. This approach can be costly, high maintenance, less flexible, and challenging when modernizing legacy software systems due to new user requirements or emerging technologies. As a result, companies seek solutions that support rapid development, continuous integration, continuous deployment, and cost-effectiveness. Recently, cloud computing has been gaining lots of attention from developers and operators. Many organizations want to adopt this emerging computing paradigm—however, the common roadblocks in migrating to cloud-based computing present enormous learning curves and management difficulties. As a result, many organizations are suffering from adopting this new computing paradigm despite its benefits. We challenge how cloud-based software reverse engineering can provide low maintenance, low cost, highly secure, highly manageable solutions for businesses. This research shares what we learned from cloud-based software reverse engineering supporting low-maintenance, low-cost, highly secure, and highly manageable solutions.

18 Sejun Song (University of Missouri Kansas City), "Apple In My Eyes (AIME): Liveness Detection for MobileSecurity Using Corneal Specular Reflections"

In this poster, we present a novel software-based face Presentation Attack Detection (PAD) method named "Apple in My Eyes (AIME)" using screen display as a challenge and corneal specular reflections as a response for authenticating the liveness against presentation. To detect face liveness, AIME creates multiple image patterns on the authentication screen as a challenge, then captures meaningful corneal specular reflection responses from user's eyes using the front camera, and analyzes the reflective pattern images using various lightweight Machine Learning (ML) techniques under a subsecond level delay (200 ms). We demonstrate that AIME can detect various attacks, including digital images displayed on the phone or tablet, printed paper images, 2D paper masks, videos, 3D silicon masks, and 3D facial models using VR. AIME liveness detection can be applied for various contactless biometric authentication accurately and efficiently without any costly extra sensors.

19 Gichan Lee (Kennesaw State University), "Interpreting Machine Learning based Code Smell Severity Prediction with Explainable AI"

Code smells are symptoms that suggest the existence of bad design and implementation choices in the source code. Although recent studies have shown good performance with predictive models that leverage machine learning (ML) techniques for code smell detection or prioritization, the rationales for each prediction performed by ML models are not well known yet. Each prediction from ML models should provide grounds related to the development context such that developers can trust the prediction results and justify the maintenance activities to be performed on the code smell. In this work, we aim to determine whether the interpretation of the model can provide explanatory power to the prediction results of ML-based code smell solutions. We trained an ML model that predicts the presence of code smell severity and interpreted the prediction results using a state-of-the-art explainable artificial intelligence technique. The results revealed that model interpretation can enable the understanding of the rationales behind each code smell severity prediction and identify the corresponding tendencies of influencing features. Based on these results, we highlight the benefits from interpretation of the model in software engineering research area.

20 Jemin Ahn (Kennesaw State University), "Rapid Web-Page Loading with Preload Scanner"

Despite being continuously advanced, conventional web page loading processes remain inefficient and prevent network devices from fully utilizing their computing capabilities. This study investigates the primary source of this inefficiency, known as script blocking, which occurs owing to browser pipelining. An advanced web page loading process is then proposed, in which the Preload Scanner in Chromium is modified such that it can identify correlations between web objects and prerequisites associated with them. Experimental testing with example websites demonstrates an improvement in overall web page load times.

21 Jiho Noh (Kennesaw State University), "Joint learning for biomedical NER and entity normalization"

This paper presents a model for biomedical named entity recognition (NER) and entity normalization (EN). More specifically, the proposed model deals with biomedical mention detection (MD), entity typing (ET), and concept mapping (or entity normalization, shorted by EN). Two different architectures are produced: The IOBHI model predicts the IOB tagging and the entity types separately, while the ONETAG model jointly predicts the BOI-type tags. Both architectures apply the pre-trained token embeddings in SciBERT, and then do the fine-tuning training by bi-LSTM-CRF and other feed-forward neural networks. State-of-the-art level performances are achieved on the MedMentions-ST21pv dataset using the F1 score. Finally, a zero-shot evaluation is applied to the model, proving that the model has zero-shot learning advantages.

22 Brian Maeng (City University of Seattle), "Coordinating an Emission-dependent Supply Chain under Cap-and-trade Regulation"

We study a two-tier supply chain problem consisting of a manufacturer and a retailer on a product under cap-and-trade regulation. The increase of sustainability effort exerted on the product results in a higher level of customer demand and decreases the carbon emission level of the manufacturer, but raises the manufacturer's cost of sustainability investment. We study the optimal decisions of the manufacturer and the retailer under both decentralized and centralized supply chains and identify the characteristics of the optimal solutions. Also, we compare various contracts and identify the one that coordinates the supply chain profit.

23 Sang Kim (Iowa State University), "Low-Latency Detection of False Data in Short Packets"

A new wave of socially useful applications based on networked sensors and autonomous devices are emerging in many fields such as the Internet of Things (IoT). Broadly speaking, these wireless applications collect data from edge devices, transport it to the controller, and analyze it to control edge devices. A core requirement of these applications is that they maintain the data integrity as the packet moves through the network. I will present a holistic paradigm of detecting and mitigating false data that are injected into short packets which are the main traffic in emerging applications. Rather than relying solely upon generic,

higher-layer cryptographic solutions (as has been the norm), a multidisciplinary approach that interplays between the cryptographic integrity check and the physical-layer integrity check is taken to assure data integrity with reduced overhead and low latency.

24 Sooyong Jang (University of Pennsylvania), "Improving Classifier Confidence using Lossy Label-Invariant Transformations"

Providing reliable model uncertainty estimates is imperative to enabling robust decision making by autonomous agents and humans alike. While recently there have been significant advances in confidence calibration for trained models, examples with poor calibration persist in most calibrated models. Consequently, multiple techniques have been proposed that leverage label-invariant transformations of the input (i.e., an input manifold) to improve worst-case confidence calibration. However, manifold-based confidence calibration techniques generally do not scale and/or require expensive retraining when applied to models with large input spaces (e.g., ImageNet). In this paper, we present the recursive lossy label-invariant calibration (ReCal) technique that leverages label-invariant transformations of the input that induce a loss of discriminatory information to recursively group (and calibrate) inputs – without requiring model retraining. We show that ReCal outperforms other calibration methods on multiple datasets, especially, on large-scale datasets such as ImageNet.

25 Jungseok Hong (University of Minnesota), "Semantically-Aware Strategies for Stereo-Visual Robotic Obstacle Avoidance"

Mobile robots in unstructured, mapless environments must rely on an obstacle avoidance module to navigate safely. The standard avoidance techniques estimate the locations of obstacles with respect to the robot but are unaware of the obstacles' identities. Consequently, the robot cannot take advantage of semantic information about obstacles when making decisions about how to navigate. We propose an obstacle avoidance module that combines visual instance segmentation with a depth map to classify and localize objects in the scene. The system avoids obstacles differentially, based on the identity of the objects: for example, the system is more cautious in response to unpredictable objects such as humans. The system can also navigate closer to harmless obstacles and ignore obstacles that pose no collision danger, enabling it to navigate more efficiently. We validate our approach in two simulated environments: one terrestrial and one underwater. Results indicate that our approach is feasible and can enable more efficient navigation strategies.

26 Sai Kosaraju (University of Nevada, Las Vegas), "HipoMap: Graphical Representation Map for WholeSlide-based Histopathological Image Analysis"

Digital pathology coupled with advanced machine learning (e.g., deep learning) has been changing the paradigm of whole-slide histopathological images (WSIs) analysis. Major applications in digital pathology using machine learning include automatic cancer classification, survival analysis, and subtyping from pathological images. While most pathological image analyses are based on patch-wise processes due to the extremely large size of histopathology images, there are several applications that predict one clinical outcome per slide (e.g., cancer classification, survival analysis). However, slide-based analysis in WSI has been seldom investigated. We propose a novel slide-based histopathology analysis framework that produces graphical representation maps, called HipoMap, for each whole slide image for efficient and effective slide-based analysis

27 Chijung Jung (University of Virginia), "Swarmbug: Debugging Configuration Bugs in Swarm Robotics"

Swarm robotics collectively solve problems that are challenging for individual robots, from environmental monitoring to entertainment. The algorithms enabling swarms allow individual robots of the swarm to plan, share, and coordinate their trajectories and tasks to achieve a common goal. Such algorithms rely on a large number of configurable parameters that can be tailored to target particular scenarios. This large configuration space, the complexity of the algorithms, and the dependencies with the robots' setup and performance make debugging and fixing swarms configuration bugs extremely challenging. This paper proposes Swarmbug, a swarm debugging system that automatically diagnoses and fixes buggy behaviors

caused by misconfiguration. The essence of Swarmbug is the novel concept called the degree of causal contribution (Dcc), which abstracts impacts of environment configurations (e.g., obstacles) to the drones in a swarm via behavior causal analysis. Swarmbug automatically generates, validates, and ranks fixes for configuration bugs. We evaluate Swarmbug on four diverse swarm algorithms. Swarmbug successfully fixes four configuration bugs in the evaluated algorithms, showing that it is generic and effective. We also conduct a real-world experiment with physical drones to show the Swarmbug's fix is effective in the real-world.

28 Sian Lee (Pennsylvania State University), "Effects of Associative Inference on People's Susceptibility to Fake News: Experimental and Empirical Evidences"

Associative inference is an adaptive process that allows people to link related information together to make novel connections which is not necessarily true. We conducted three online human-subject experiments (N = 3,015) investigating the effects of associative inference on people's susceptibility to fake news on social media. In each experiment, participants completed recognition and perceived accuracy rating tasks for the snippets of news articles in the tweet format across two phases. In Phase 1, participants viewed real-news pairs only. In Phase 2, participants viewed both real and fake news. Critically, participants saw exactly the same real and fake news in Phase 2, but we varied the association between real-news pairs at Phase 1 and fake news at Phase 2 across three levels (control, association only, and associative inference). Both recognition and perceived accuracy results showed that participants were more susceptible to fake news with associative inference. Also, we identified associative inference cases in the wild by mining a Twitter dataset. Specifically, we crawled 1.7 M of Tweets from Twitter users who "agree" (e.g., tweets with agreeing terms, re-tweet, or like) on some fake news. Next, using a modified Apriori algorithm in associative rule mining, we performed data-driven analysis to see if there exist real Twitter users whose timeline feed contains the associative inference cases, and found an evidence that associative inference in the past may have influenced the acceptance of fake news afterward, though the causal influence of associative inference in the wild still remains as our future work. Altogether, our work not only suggests that associative inference, i.e., constructive process of memory, can be one basis for people's susceptibility to fake news, but also implies the potential actual influence of associative inference in the wild.

29 Sian Lee (Pennsylvania State University), "Optimal auction order with two different items in sequential Vickrey auctions with reserve price"

Algorithmic auction theory is one of the major subfields of algorithmic game theory (AGT). I develop an auction algorithm which verifies that the auction order for sequential sealed-bid second-price auction (i.e., Vickrey auction) with reserve prices has an impact on the seller's expected payoff in a specific setting. In addition, the auction algorithm shows which case gives the higher expected payoff to the seller and how it is changed according to the change of the distribution of the items at Perfect Bayesian Nash Equilibrium (PBNE). Specifically, I assume a model with the two different items (a, b), two bidders, and two cases (Case A, Case B). Furthermore, it is assumed that items "a" and "b" have a two-point distribution with $V_a \in \{10, 11\}$ and $V_b \in \{10, 12\}$. Note that the expected value of item "b" is higher than item "a." Thus, item "b" represents the high-quality item compared to item "a." Moreover, a high value for each item ($V_a=11, V_b=12$) is realized with probability "p," and a low value for each item ($V_a=10, V_b=10$) is realized with probability "1 - p." In case A, the seller sells the item "a" first, and in case B, the seller sells the item "b" first. Equilibria are found for each case, and the seller's expected payoff for each case at equilibrium is compared. In summary, if "p" is low, case B which sells the high-quality item (item "b") in the first auction through competition, gives the seller a higher expected payoff than case A. In contrast, if "p" is sufficiently high, then case A which sells the high-quality item (item "b") in the second auction and extracts all the surplus from the high-type bidder through the reserve price, gives the seller a higher expected payoff than case B.

30 Kijoon Bang (University of California, Los Angeles), "Bayesian reaction optimization in solution based thermal reaction toward single crystalline perovskite semiconductor"

Reaction optimization is fundamental to material growth of novel semiconductor, which is prerequisite for single crystal generation process. However, due to extremely high degrees of freedom in solution based synthetic process of halide perovskite material class, it takes extensive amount of costs and time to figure out key parameters. Conventionally, design of experiments (DOE) is commonly used in evaluating and predicting experimental conditions, which are heavily confined to precedent experimental results. This confinement requires some variables be constant for cost effective optimization, which largely limited in real-world

examples like model-driven approaches. Here, in this work, Bayesian optimization, which is based on uncertainty guided response surface method is used explain synthetic process of single crystalline material growth process. Bayesian optimization algorithms are exploitable in searching for diverse spaces spanned by arbitrary reaction domains so that many layers of parallel experiments can be explored. For the synthesis of single perovskite crystals, one of the reaction domains provide extensive surface tension of the solution with the lowest Gibbs energy. The length of time is also assessment crucial factor in the final stage of optimization. Finally, yield of the crystal products are also investigated compare with the results reported earlier.

31 Bumsu Kim (University of California, Los Angeles), "Curvature-Aware Derivative Free Optimization"

We propose a new line-search method, coined Curvature-Aware Random Search (CARS), for derivative-free optimization. CARS exploits approximate curvature information to estimate the optimal step-size given a search direction. We prove that for strongly convex objective functions, CARS converges linearly if the search direction is drawn from a distribution satisfying very mild conditions. We also explore a variant, CARS-NQ, which uses Numerical Quadrature instead of a Monte Carlo method when approximating curvature along the search direction. We show CARS-NQ is effective on highly non-convex problems of the form $f=f_{cvx}+f_{osc}$ where f_{cvx} is strongly convex and f_{osc} is rapidly oscillating. Experimental results show that CARS and CARS-NQ match or exceed the state-of-the-arts on benchmark problem sets.

32 Junho Hong (University of Michigan - Dearborn), "Cybersecurity of EV charging station"

The use of industrial control systems (ICS) for controlling critical components in a power grid is becoming more prevalent. Additionally, asset monitoring devices employing sensors for reporting voltage, current, temperature, and other measurements are becoming more common for power system operations. However, due to the high penetration of Information and Communications Technology (ICT), Supervisory Control And Data Acquisition (SCADA) systems are interconnected with one another, resulting in higher vulnerability with respect to cyber intrusions. Cyberattacks on critical infrastructures are evolving, and their patterns are diversifying, particularly for energy delivery systems. Cyberattacks can damage physical systems by compromising their ICT infrastructure and gaining access to the controls and monitoring of energy delivery systems, without a requiring a physical attack. In this presentation, state-of-the-art cyberattack detection, impact analysis and mitigation methods will be discussed together with their practical implementations for extreme fast charging station. New intelligent of cyber security design, operation, and metrics will be important for resiliency of future energy delivery systems.

Sponsors

Gold Sponsors

Samsung Research

SAMSUNG Research

KSEA (Korean-American Scientists and Engineers Association)



Silver Sponsors

Korean-U.S. Science Cooperation Center, Co-organizer



LG Electronics



Bronze Sponsors

Naver



Bobarobo



Scholarship Support

Moon-Jung Chung Family

<https://research.samsung.com>

Shape the Future with Innovation and Intelligence

Samsung Research is the advanced research and development (R&D) hub of Samsung's Consumer Electronics (CE) Division and IT & Mobile Communications (IM) Division. Samsung Research leads the development of the future technologies with more than 10,000 researchers and developers working in overseas R&D centers.

Under the vision of "Shape the Future with Innovation and Intelligence", it is actively conducting research and development (R&D) to identify new future growth areas and secure advanced technologies for its products & services to create new value and improve people's lives.

Core research themes at Samsung Research include artificial intelligence (AI), data intelligence, next-generation communications, robot, Tizen, life care & new experiences, next-generation media, and security. In particular, it is expanding its research scope to new promising fields to realize a new lifestyle based on AI technologies.

The hub is working in collaboration with 14 overseas R&D centers in 12 countries worldwide and 7 global AI centers to secure innovative technologies and enhance its global R&D capacity. It is equally maximizing technological cooperation through active open innovation with distinguished universities, research institutes, and partner companies globally that have the world's best technologies.

Samsung Research promises to continue working hard to become a global top research institute that creates new values for the future through ceaseless innovation and intelligence.

Samsung Research







The Korea-U.S. Science Cooperation Center (KUSCO) is a non-profit organization established in Vienna, Virginia, founded in 1997, to accomplish two major missions: to enhance cooperative efforts in S&T between Korea and the U.S. and to support Korean - American scientists & engineers in the United States.



VISION & MISSION

The vision of KUSCO is *"Advancement of the S&T collaborations between Korea and the U.S."*

To realize its vision, KUSCO pursues the mission of becoming a premium center for S&T cooperation and Exchanges by pursuing three goals as follows:

- To strengthen S&T cooperation portfolio between Korea and the U.S.
- To broaden global exchange programs
- To support and leverage Korean-American scientists and engineers



MAJOR ACTIVITIES

- Enhance scientific and technological development of both Korea and the U.S.
- Support mutual cooperation and initiate joint programs with U.S. and Korean scientific and engineering societies, academic universities, and other institutions
- Assess significant trends in scientific research and technological developments affecting Korea and the U.S.
- Assist young Korean-American and other scientists in developing and maintaining networks addressing matters of scientific and technical interest to the two countries

