IEEE BigData 2024 Program Schedule
Washington, DC, USA
Conference time zone: US Eastern Standard Time

The link to download the BigData 2024 conference proceedings for BigData24 and access the pre recorded videos is:

Meeting Room

Floor	Room
Lobby	CONGRESSIONAL A
Lobby	CONGRESSIONAL B
Lobby	CONGRESSIONAL C
Lobby	CONGRESSIONAL D
Lobby	CONGRESSIONAL E
Lobby	CAPITOL A
Lobby	CAPITOL B
Ballroom	BUNKER HILL
Ballroom	LEXINGTON
Ballroom	CONCORD
Ballroom	COLUMBIA A
Ballroom	COLUMBIA B
Ballroom	COLUMBIA C
Ballroom	REGENCY A
Ballroom	REGENCY B
Ballroom	REGENCY C
Ballroom	REGENCY D
2nd floor	YELLOWSTONE
2nd floor	EVERGLADES
2nd floor	GRAND TETON
2nd floor	BRYCE
2nd floor	YOSEMITE
2nd floor	GLACIER
2nd floor	SEQUOIA
2nd floor	OLYMPIC

Sponsored by

Platinum Sponsor:

National Science Foundation

Gold Sponsor:

VIRGINIA TECH

IEEE*DataPort*™

Silver Sponsor:

ankura

IEEE Big Data 2024 Program Schedule

Washington DC, USA
December 15 - December 18, 2024

Keynote Lecture: **60 minutes** (50-min talk + 10-min Q/A)
Main conf. regular: **25 mins** (20-minute talk + 5-min Q/A)
Main conf. short: **15 minutes** (11-min talk + 4-min Q/A)

Registration: Saturday, Dec 14 16:00-18:00 PM
(Location: Regency Foyer)

Day 1: Sunday, December 15, 2024			
Time	**Location**	**Workshop, Tutorial, Special Symposium, Special Session**	**Chairs**
7:30-17:00	Regency Foyer	**Registration**	Kunpeng Liu, Dongjin Song
10:00-10:30	Columbia Foyer	**Coffee Break**	
12:30-14:00	**Lunch (on your own)**		
16:00-16:30	Columbia Foyer	**Coffee Break**	
8:30-12:30	Congressional - A	Workshop #15: The 4th International Workshop on Multi-Modal Medical Data Analysis	Bo Zhang
14:00-18:00	Congressional - A	Workshop #48: DCAI: The 3rd International Workshop on Data-Centric AI	Yanjie Fu
8:30-12:30	Congressional - B	Workshop #51: Big Data and Artificial Intelligence for Personalized Health Care	Ryan Benton
9:00-17:00	Congressional - C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
9:00-17:00	Congressional - D	Workshop #13: The 7th Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen
8:30-12:30	Capitol Room - A	Workshop #17: The 3rd Workshop on Big Data Analytic in Healthcare	Alex Kuo and Matteo Mantovani
14:00-18:00	Capitol Room - A	Workshop #20: Next-Generation Big Graph Learning and Analytics (Next-Gen-BiGLA)	Xiaofei Zhang

9:00-17:00	Capitol Room - B	Workshop #23: The 5th International Workshop on Big Data & AI Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD) 2024	Sangkeun Lee
8:30-12:30	Concord	Workshop #36: Second International Workshop for securing IoT systems using BigData	Cristian Bucur
14:00-18:00	Concord	Workshop #39: The Seventh Workshop on Big Data for Economic and Business Forecasting	Wei Shang
9:00-17:00	Columbia - A	Workshop #7: 2nd Workshop on AI Music Generation with AI Music Competition (AIMG 2024)	Callie Liao
9:00-17:00	Columbia - B	Workshop #38: Workshop on Big Data and AI for Healthcare	Ping Chen
9:00-17:00	Columbia - C	Workshop #9: 4th Workshop on Knowledge Graphs and Big Data In conjunction with the IEEE Big Data 2024	Yuan An
9:00-17:00	Regency - B	High School and Undergrad Symposium	Xuan Wang & Yanjie Fu
9:00-17:00	Regency - C	Special Session: Federated Learning on Big Data	Francesco Piccialli
9:00-17:00	Regency - D	REU Symposium	Xuechen Zhang, Xinhui Zhao, and Xiaokun Yang
9:00-17:00	Yellowstone	Workshop #2: 7th Annual Workshop on Cyber Threat Intelligence and Hunting (CyberHunt 2024)	Vasileios Mavroeidis
9:00-17:00	Everglades	Workshop #46: 7th Workshop on Big Data for CyberSecurity (BigCyber-2024)	Lavanya Elluri
9:00-17:00	Grand Teton	Workshop #3: 2nd workshop on Big Data Applications for Fight against Crime and Terrorism (BDA4FCT)	Georgios Stavropoulos
9:00-17:00	Bryce	Workshop #5: The 9th International Workshop on Application of Big Data for Computational Social Science (ABCSS2024)	Isamu Okada
8:30-12:30	Yosemite	Workshop #4: 8th International Workshop on Applications of Big Data Methods and Technology in the Transport Industry	John Easton

14:00-18:00	Yosemite	Workshop #47: 9th Workshop on Advances in High Dimensional (AdHD) Big Data	Aristidis Vrahatis
9:00-17:00	Sequoia	Special Session: Intelligent Data Mining	Uraz Yavanoglu

Day 2: Monday, December 16, 2024			
8:00-17:30	**Registration**		
8:00-8:30	**Poster Setup**		
8:20-9:00	**Opening Ceremony** Conference Chairs: Chang-Tien Lu (Virginia Tech) Fusheng Wang (Stony Brook U.) PC Chairs: Wei Ding (U. of Massachusetts, Boston) Liping Di (George Mason U.) Ricardo Baeza-Yates (Northeastern U.) I&G Chairs: Kesheng Wu, (Lawrence Berkeley National Lab) Raghu Nambiar (AMD) Luke Jun Huan (Amazon) Steering Committee Chair: Xiaohua (Tony) Hu (Drexel U.) Location: Regency Ballroom		
9:00-10:00	**Keynote Speech I** Dr. Haixun Wang, Instacart, USA **Generative Information Retrieval and E-commerce** Location: Regency Ballroom		
10:00-10:30	Columbia Foyer	**Coffee Break**	
10:00-10:30	Regency Foyer Columbia Foyer	**Poster Display** (Session Chairs:Kaiqun Fu, Yufei Tang)	
Time	**Location**	**Session/Workshop**	**Session Chair**
10:30-12:30	Capitol A	D1S1: Foundation Models for Big Data	
10:30-12:30	Capitol B	D1S2: Big Data Learning and Analytics I	
10:30-12:30	Columbia A	D1S3: Big Data Infrastructure I	
10:30-12:30	Columbia C	D1S4: Big Data Science and Foundations I	
10:30-12:30	Congressional A	D1S5: Big Data Management I	
10:30-12:30	Congressional B	D1S6: Big Data Search and Mining I	
10:30-11:30	Columbia B	North American Student Challenge	

		Competition	
9:00-17:00	Congressional D	Workshop #42: From Theory to Practice: Workshop on Large Language and Foundation Models	Tania Cerquitelli
10:30-12:30	Yellowstone	Challenge Cup: Predicting Chess Puzzle Difficulty	Dominik Ślęzak
14:00-18:00	Everglades	Workshop #34: The 8th Annual Workshop on Applications of Artificial Intelligence in the Legal Industry	Han Qin
8:30-12:30	Yosemite	Workshop #8: 4th IEEE Big Data Workshop on Multimodal AI (MMAI 2024)	Lindi Liao
14:00-18:00	Grand Teton	Workshop #6: 10th Workshop on Methods to Improve Big Data Science Projects (MIDP-2024)	Jeffrey Saltz
8:30-12:30	Bryce	Joint Session Workshop #14: 2nd WS on Large-scale Data Utilization in Economics of Information and Management Sciences: Theory, Computation, and Experiment Special Session: Understanding New Markets by Data Science, Social Science, and Economics	Teruaki Hayashi Naoki Watanabe Hiroki Sakaji
10:30-12:30	Concord	I&G-1: Generative AI	Kesheng Wu
10:30-12:30	Glacier	Special Session: Federated Learning on Big Data	Francesco Piccialli
10:30-12:30	Congressional C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
10:30-12:30	Sequoia	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
12:30-14:00	**Lunch**		
14:30-16:00	Regency Ballroom	Panel: Big Data and AI: How they benefit each other	Bhavani Thuraisingham
14:00-16:00	Capitol A	D1S7:Data Ecosystem	
14:00-16:00	Capitol B	D1S8: Big Data Learning and Analytics II	
14:00-16:00	Columbia A	D1S9: Big Data Learning and Analytics III	

14:00-16:00	Columbia B	D1S10: Deep Learning I	
14:00-16:00	Congressional A	D1S11: Machine Learning Algorithms I	
14:00-16:00	Congressional B	D1S12: Deep Learning II	
14:00-16:00	Concord	I&G-2: Big data Infra & Multi-modality	Kesheng Wu
14:00-16:00	Columbia C	I&G-3: Modeling	Kesheng Wu
14:00-16:00	Bryce	Joint Session Workshop #14: 2nd WS on Large-scale Data Utilization in Economics of Information and Management Sciences: Theory, Computation, and Experiment Special Session: Understanding New Markets by Data Science, Social Science, and Economics	Teruaki Hayashi Naoki Watanabe Hiroki Sakaji
14:00-16:00	Glacier	Special Session: Federated Learning on Big Data	Francesco Piccialli
14:00-16:00	Congressional C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
14:00-16:00	Sequoia	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
16:00-16:30	Columbia Foyer	**Coffee Break**	
16:00-16:30	Regency Foyer Columbia Foyer	**Poster Display** (Session Chairs: Kaiqun Fu, Yufei Tang)	
16:30-18:00	Capitol A	D1S13: Big Data Applications I	
16:30-18:00	Capitol B	D1S14: Big Data Applications II	
16:30-18:00	Columbia B	D1S15: Big Data Analytics I	
16:30-18:00	Congressional A	D1S16: Big Data Analytics II	
16:30-18:00	Congressional B	D1S17: Big Data Analytics III	
16:30-18:00	Bryce	Joint Session Workshop #14: 2nd WS on Large-scale Data Utilization in Economics of Information and Management Sciences: Theory, Computation, and Experiment Special Session: Understanding New	Teruaki Hayashi Naoki Watanabe Hiroki Sakaji

Time	Location	Session/Workshop	Session Chair
		Markets by Data Science, Social Science, and Economics	
16:30-18:00	Glacier	Special Session: Federated Learning on Big Data	Francesco Piccialli
16:30-18:00	Congressional C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
16:30-18:00	Sequoia	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
19:00-21:00		OC Business Meeting	

Day 3: Tuesday, December 17, 2024			
8:30-17:00	**Registration (Regency Foyer)**		
8:30-9:00	**PhD Forum Poster Setup**		
9:00-10:00	**Keynote Speech II** Prof. Cornelia Caragea, USA and U. of Illinois Chicago, USA **Improving Semi-Supervised Learning with Pseudo-Margins** Location: Regency Ballroom		
10:00-10:30	Columbia Foyer	**Coffee Break**	
10:00-10:30	Regency Foyer Columbia Foyer	**PhD Forum Poster Display** (Session Chairs: Feng Chen, Shuhan Yuan)	
Time	**Location**	**Session/Workshop**	**Session Chair**
10:30-12:00	Regency Ballroom	Panel B: Big Data and AI – Shaping the Next Frontier of Data-Driven Intelligence	Duoduo (Lindi) Liao
10:30-12:30	Capitol A	D2S1: Foundation Models and Analytics	
10:30-12:30	Capitol B	D2S2: Learning from Complex Data	
10:30-12:30	Columbia A	D2S3: Big Data Infrastructure II	
10:30-12:30	Columbia B	D2S4: Big Data Science and Foundations II	
10:30-12:30	Congressional A	D2S5: Big Data Search and Mining II	
10:30-12:30	Congressional B	D2S6: Big Data Applications IV	
8:15-17:00	Congressional D	Workshop #1: 9th Computational Archival Science (CAS) Workshop	Mark Hedges

10:30-12:30	Columbia C	I&G-4: Business and Energy	Kesheng Wu
10:30-12:30	Yosemite	Special Session: Synergizing Mobility Data for Creating and Discovering Valuable Places	Yukio Ohsawa
10:30-12:30	Grand Teton	Special Session: Social Cognitive Computing in Digital Education and Learning	Jerry Chun-Wei Lin
10:30-12:30	Glacier	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
10:30-11:30	Concord	PhD Forum: Paper Session	Feng Chen, Shuhan Yuan
11:30-12:30	Concord	PhD Forum Panel: Beyond the Dissertation: Insights into Academic and Industry Careers	Feng Chen
12:30-14:00		**Lunch**	
14:30-16:00	Regency Ballroom	Panel: Federal Big Data Initiatives and Activities	Liping Di
14:00-16:00	Capitol A	D2S7: Multimodal and Scalable Big Data	
14:00-16:00	Capitol B	D2S8: Link and Graph Mining I	
14:00-16:00	Columbia A	D2S9: Machine Learning Algorithms II	
14:00-16:00	Yosemite	Special Session: Synergizing Mobility Data for Creating and Discovering Valuable Places	Yukio Ohsawa
14:00-16:00	Grand Teton	Special Session: Data-Driven Designation and Implementation of Automated Guided Vehicles	Jerry Chun-Wei Lin
14:00-16:00	Glacier	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
14:00-17:30	Concord	PhD Forum: Paper Session	Feng Chen, Shuhan Yuan
14:00-17:00	Yellowstone	Workshop #10: International Workshop on Data Engineering and Modeling for AI (DEMAI)	Carlos Ordonez
16:00-16:30	Columbia Foyer	**Coffee Break**	
16:00-16:30	Regency Foyer Columbia Foyer	**PhD Forum Poster Display** (Session Chairs: Feng Chen, Shuhan Yuan)	

16:30-17:00	Regency Ballroom	**Sponsor Session: IEEE Dataport**
17:30-18:30		**Walk to the National Museum of American History** 4th St. & Independence Ave. SW Washington, DC 2056
18:30-21:00	National Museum of the American Indian	**Banquet** Award Ceremony Social Program

Day 4: Wednesday, December 18, 2024			
8:30-17:00	**Registration (Regency Foyer)**		
9:00-10:00	**Keynote Speech III** Prof. Alexander "Sasha" Rush, Cornell Tech, USA **Designing Text Embeddings for the Future** Location: Regency Ballroom		
10:00-10:30	Columbia Foyer	**Coffee Break**	
Time	**Location**	**Session/Workshop**	**Session Chair**
10:30-12:30	Capitol A	D3S1: Link and Graph Mining II	
10:30-12:30	Capitol B	D3S2: Machine Learning Algorithms III	
10:30-12:30	Columbia A	I&G-5: Recommender Systems	Kesheng Wu
10:30-12:30	Columbia B	I&G-6: AI4Science	Kesheng Wu
8:30-12:30	Columbia C	Workshop #24: 2nd International Workshop on Big Data Analytics with Artificial Intelligence for Climate Change	Rajiv Misra
10:30-16:00	Congressional A	Special Session: HealthCare Data	Ozgun Pinarer
10:30-12:30	Congressional B	Special Session: Dataspaces and DFFT (Data Free Flow with Trust)	Noboru Koshizuka
10:30-12:30	Yosemite	Special Session: Synergizing Mobility Data for Creating and Discovering Valuable Places	Yukio Ohsawa
12:30-14:00	**Lunch**		
13:00-13:30	Regency Ballroom	**Closing Ceremony**	

Main Conference Paper Sessions

ID	Title	Authors	Type (Red indicates an online video presentation, not played in the conference room. Access it via the prerecorded video link in the conference program.)
	D1S4: Big Data Science and Foundations I		
BigD 400	Luckiness Normalized Maximum Likelihood-based Change Detection for High-dimensional Graphical Models with Missing Data	Zhongyi Que, Linchuan Xu, and Kenji Yamanishi	Regular
BigD 587	Scalable Sampling for High Utility Patterns	Lamine Diop and Marc Plantevit	Short
BigD 419	A Planner for Scalable Tensor Programs	Tanvir Ahmed Khan and Leonidas Fegaras	Regular
BigD 462	A Full-History Network Dataset for BTC Asset Decentralization Profiling	Ling Cheng, Qian Shao, Fengzhu Zeng, and Feida Zhu	Regular
BigD 592	Synthetic Data Generation for Photovoltaic Systems: Enhancing Accuracy with the Huld Model	Ali AGHAZADEH ARDEBILI, Andreas LIVERA, Antonella Longo, George E. GEORGHIOUS, and Antonio FICARELLA	Short
BigD 638	An Efficient Single-pass Online Computation of Higher-Order Bivariate Statistics	Anton Kochepasov and Ilya Stupakov	Short
BigD 344	Faster Sampling Algorithms for Polytopes with Small Treewidth	Yekun Ke, Xiaoyu Li, Zhao Song, and Tianyi Zhou	Regular
BigD 310	DyMGCN: Dynamic Multi-Graph Convolution Networks for Spatio-Temporal Forecasting	Hexiang Liu, Qilong Han, Jingyu Sheng, Dan Lu, Shanshan Sui, and Hui Sun	Short
	D2S4: Big Data Science and Foundations II		
BigD 665	RPS: A Generic Reservoir Patterns Sampler	Lamine Diop, Marc Plantevit, and Arnaud Soulet	Short
BigD 786	Fairness in Monotone k-submodular Maximization: Algorithms and Applications	Yanhui Zhu, Samik Basu, and A. Pavan	Short

BigD 844	Minimizing Regret in Social Media Advertisement	Poonam Sharma, Dildar Ali, and Suman Banerjee	Short
BigD 778	Nuclear Pore Segmentation in 3D FIB-SEM Images with Dynamic Cyclical Data Augmentation	Chongyu He, Zhiwu Xie, Yinlin Chen, and Edward Fox	Short
BigD 793	Firecast Zigzag Convolutional Network for Wildfire Prediction	Yuzhou Chen, Joel Castillo, Huikyo Lee, and Yulia R. Gel	Short
BigD 693	Disentangled Conditional Variational Autoencoder for Unsupervised Anomaly Detection	Asif Ahmed Neloy and Maxime Turgeon	Short
BigD 593	Proposing a Low-Rank Approximation Method with Mathematical Guarantees for High-Dimensional Tensor Data	Hiroki Hasegawa, Kazuyoshi Yata, Yukihiko Okada, and Jun Kunimatsu	Regular
BigD 439	Learning with Sparsely Permuted Data: A Robust Bayesian Approach	Abhisek Chakraborty and Saptati Datta	Regular
BigD 546	ICL: An Incentivized Collaborative Learning Framework	Xinran Wang, Qi Le, Ahmad Khan, Jie Ding, and Ali Anwar	Regular
BigD 741	CRTRE: Causal Rule Generation with Target Trial Emulation Framework	Junda Wang, Weijian Li, Han Wang, Hanjia Lyu, Caroline Thirukumaran, Addisu Mesfin, Hong Yu, and Jiebo Luo	Short
BigD 518	Zero-th Order Algorithm for Softmax Attention Optimization	Yichuan Deng, Zhihang Li, Sridhar Mahadevan, and Zhao Song	Regular
BigD 821	[Vision Paper] Visual Summary Thought of Large Vision-Language Models for Multimodal Recommendation	Yuqing Liu, Yu Wang, Yuwei Cao, Lichao Sun, and Philip Yu	Short
BigD 887	Optimal Scheduling Algorithms for Cost-Effective Bandwidth Reservation in HPNs	Liudong Zuo, Pan Lai, and Zhong Chen	Short
BigD 812	Unsupervised Domain Adaptation for Entity Blocking Leveraging Large Language Models	Yaoshu Wang and Mengyi Yan	Short
BigD 709	Evaluating Blocking Biases in Entity Matching	Mohammad Hossein Moslemi, Harini Balamurugan, and Mostafa Milani	Regular
	D1S3: Big Data Infrastructure I		
BigD 240	Stochastic Large-scale Machine Learning Algorithms with Distributed Features and Observations	Biyi Fang, Diego Klabjan, and Truong Vo	Regular
BigD 255	MPI Allgather Utilizing CXL Shared Memory Pool in Multi-Node Computing Systems	Hooyoung Ahn, Seonyoung Kim, Yoomi Park, Woojong Han, Shinyoung Ahn, Tu Tran, Bharath Ramesh, Hari	Short

		Subramoni, and Dhabaleswar K. Panda	
BigD 313	Large-Scale Knowledge Graph Embeddings in Apache Spark	Bedirhan Gergin and Charalampos Chelmis	Regular
BigD 329	Key Compression Limits for k-Minimum Value Sketches	Charlie Dickens, Eric Bax, and Alexander Saydakov	Short
BigD 314	CacheIt: Application-Agnostic Dynamic Caching for Big Data Analytics	Dat Nguyen, Muhammad Rafid, Nathanael Santoso, and Khanh Nguyen	Regular
BigD 373	Optimizing Privacy While Limiting Information Loss in Distributed Data Anonymization	Arsème Vadèle Djeufack Nanfack, Gaël Le Mahec, and Gilles Dequen	Short
	D2S3: Big Data Infrastructure II		
BigD 320	Efficient Data-parallel Distributed DNN Training for Big Dataset under Heterogeneous GPU Cluster	Shinyoung Ahn, Sookwang Lee, Hyeonseong Choi, and Jaehyun Lee	Regular
BigD 569	DBRepo: A Data Repository System for Research Data in Databases	Martin Weise and Andreas Rauber	Regular
BigD 423	Common Pitfalls with Data Deduplication Parameters and Metrics	Luke Schultz, Owen Randall, and Paul Lu	Regular
BigD 735	Distributed Data Analysis Workflow System across Multiple Data Hubs	Siwoon Son, Seong-Hwan Kim, and Heesun Won	Short
BigD 586	Durin: CPU-FPGA Heterogeneous Platform for Scalable Low-Dimensional Data Clustering	Se-Min Lim, Esmerald Aliaj, and Sang-Woo Jun	Short
BigD 611	An Overview of the Data-Loader Landscape: Comparative Performance Analysis	Iason Ofeidis, Diego Kiedanski, and Leandros Tassiulas	Short
BigD 547	Towards Scalable Quantum Repeater Networks	Connor Howe, Mohsin Aziz, and Ali Anwar	Short
BigD 461	Improving Data Science Applications with a Visual Cross-Platform Execution Environment	Lucas Miguel Ponce and Dorgival Guedes	Regular
	D1S5: Big Data Management I		
BigD 447	Towards True Multi-interest Recommendation: Enhanced Scheme for Balanced Interest Training	Jaeri Lee, Jeongin Yun, and U Kang	Regular
BigD 823	Uncertainty-Aware Influence Maximization: Enhancing Propagation in Competitive Social Networks with Subjective Logic	Qi Zhang, Lance Kaplan, Audun Jøsang, Dong Hyun Jeong, Feng Chen, and Jin-Hee Cho	Short
BigD 722	Scalable Big Spatial Data Processing with SQL Query Compilation and Distributed Morsel-driven Parallelism	Rahul Sahni, Xiaozheng Zhang, Sudip Chatterjee, and Suprio Ray	Regular
BigD 508	Session-based News Recommendation Using Cohesive Patterns	Mozhgan Karimi, Len Feremans, Boris Cule, and Bart Goethals	Short

BigD 695	$\ell_{1,2}$-Norm and CUR Decomposition based Sparse Online Active Learning for Data Streams with Streaming Features	Zhong Chen, Yi He, Di Wu, Liudong Zuo, Keren Li, Wenbin Zhang, and Zhiqiang Deng	Regular
BigD 644	Synthesizing Accurate Relational Data under Differential Privacy	Antheas Kapenekakis, Daniele Dell'Aglio, Charles Vesteghem, Laurids Poulsen, Martin Bøgsted, Minos Garofalakis, and Katja Hose	Short
BigD 265	Personalized Multi-task Training for Recommender System	Liangwei Yang, Zhiwei Liu, Jianguo Zhang, Rithesh Murthy, Shelby Heinecke, Huan Wang, Caiming Xiong, and Philip S. Yu	Regular
BigD 589	Vision Paper: FAIR Digital Objects for Realising Globally Aligned Data Spaces	Nicolas Blumenröhr, Philipp-Joachim Ost, Felix Kraus, and Achim Streit	Regular
BigD 370	Understanding the Intrinsic Characteristics of Spatial Partitioning in Distributed Spatial Join	Zhuohan Xu, Dejun Teng, Zhaohui Peng, and Fusheng Wang	Regular
BigD 312	Leveraging LLMs for Fair Data Labeling and Validation in Crowdsourcing Environments [Vision paper]	Ibrahim Mohamed Serouis and Florence Sèdes	Short
BigD 433	Causally Informed Factorization Machines	Mao-Lin Li, K. Selçuk Candan, and Maria Luisa Sapino	Short
BigD 420	Transferable Sequential Recommendation via Vector Quantized Meta Learning	Zhenrui Yue, Huimin Zeng, Yang Zhang, Julian McAuley, and Dong Wang	Short
	D1S6: Big Data Search and Mining I		
BigD 267	Multi-Modality Transformer for E-Commerce: Inferring User Purchase Intention to Bridge the Query-Product Gap	Srivatsa Mallapragada, Ying Xie, Varsha Rani Chawan, Zeyad Hailat, and Yuanbo Wang	Regular
BigD 401	SHRINK: Data Compression by Semantic Extraction and Residuals Encoding	Guoyou Sun, Panagiotis Karras, and Qi Zhang	Regular
BigD 487	On Modeling Adaptive Index Management as Adversarial Search	Gajendra Doniparthi, Tim Otto, and Stefan Dessloch	Regular
BigD 317	TensorSearch: Parallel Similarity Search on Tensors	Bin Dong, Avinash Nayak, Kesheng Wu, Verónica Rodríguez Tribaldos, Jonathan Ajo-Franklin, Qile Zhang, Suren Byna, Fan Guo, Patrick Dobson, and Alexander Sim	Regular

BigD 383	GNN-Based Persistent K-core Community Search in Temporal Graphs	Zongli Jiang, Yirui Tan, Guoxin Chen, Fangda Guo, Jinli Zhang, and Xiaolu Bai	Regular
BigD 225	Closer through commonality: Enhancing hypergraph contrastive learning with shared groups	Daeyoung Roh, Donghee Han, Daehee Kim, Keejun Han, and Mun Yong Yi	Regular
BigD 304	AGO-FT: An adaptive guided oversampling based on fast space division and trustworthy sampling space for imbalanced noisy datasets	Yi Deng, Min Wu, and Yan Ma	Regular
	D2S5: Big Data Search and Mining II		
BigD 427	CANDE: A Lightweight Locality-Sensitive Hashing Add-on for Candidate-Based Distribution Estimation	Jingfan Meng, Huayi Wang, Kexin Rong, and Jun Xu	Regular
BigD 658	Optimizing Diverse Information Exposure in Social Graphs	Jonathan Colin and Silviu Maniu	Regular
BigD 539	DyGCL: Dynamic Graph Contrastive Learning For Event Prediction	Muhammad Ifte Khairul Islam, Khaled Mohammed Saifuddin, Tanvir Hossain, and Esra Akbas	Regular
BigD 705	Fast Distributed Memory Parallel Algorithms for Finding Connected Components in Large Graphs	Maleq Khan and Sharon Boddu	Regular
BigD 499	Retrieving Top-k Hyperedge Triplets: Models and Applications	Jason Niu, Ilya Amburg, Sinan Aksoy, and Ahmet Erdem Sariyuce	Regular
BigD 617	Parallel Motif-Based Community Detection	Tianyi Chen and Charalampos Tsourakakis	Regular
BigD 642	Zero-Shot Relational Learning for Multimodal Knowledge Graphs	Rui Cai, Shichao Pei, and Xiangliang Zhang	Regular
BigD 691	Automated Polynomial Filter Learning for Graph Neural Networks	Wendi Yu, Zhichao Hou, and Xiaorui Liu	Regular
	D2S7: Multimodal and Scalable Big Data		
BigD 836	A machine-learning pipeline for merging and georeferencing very large archives of historical aerial photographs	Eugenio Noda, Luna Huang, Trinetta Chong, Shruti Jain, Andreas Madestam, Anna Tompsett, Hannah Druckenmiller, and Solomon Hsiang	Regular
BigD 759	Custom Accessors: Enabling Scalable Data Ingestion, (Re-)Organization, and Analysis on Distributed Systems	Vito Giovanni Castellana, Burcu Mutlu, Ian Di Dio Lavore, Jesun Firoz,	Regular

			Katherine Wolf, Marco Minutoli, and John Feo	
BigD 711	Efficient Point-to-Subspace ANNS in Manhattan and Lp Space by LSH Pruning		Jingfan Meng, Huayi Wang, and Jun Xu	Regular
BigD 779	QualityNet: Error-bounded Lossy Compression Quality Prediction via Deep Surrogate		Khondoker Mirazul Mumenin, Dong Dai, Jinzhen Wang, and Sheng Di	Regular
BigD 811	A Study of Data-Path Bugs in PyTorch with a Focus on Memory Management		Rubayet Rahman Rongon, Chen Cao, and Xuechen Zhang	Regular
BigD 867	Evaluating Text Summarization Techniques and Factual Consistency with Language Models		Md Moinul Islam, Usman Muhammad, and Mourad Oussalah	Short
BigD 879	SAFENet: Towards a Robust Suicide Assessment Using Selective Prediction Framework		Usman Naseem, Surendrabikram Thapa, Siddhant Shah, Mohammad Salman, Qi Zhang, Junaid Rashid, Liang Hu, and Imran Razzak	Regular
BigD 767	Fast Sentence Classification using Word Co-occurrence Graphs		Ashirbad Mishra, Shad Kirmani, and Kamesh Madduri	Regular
	D2S8: Link and Graph Mining I			
BigD 394	A Novel Reverse Random Hyperplane Projection Scheme and Its Effect on Mining Sensor Streams		Antonios Skevis, George Klioumis, and Nikos Giatrakos	Short
BigD 770	ID and Graph View Contrastive Learning with Multi-View Attention Fusion for Sequential Recommendation		Xiaofan Zhou and Kyumin Lee	Regular
BigD 410	Monarch: Distributed Butterfly Counting for Large-scale Bipartite Graph		Yuxin Tang, Mangesh Bendre, and Mahashweta Das	Short
BigD 421	Enhancing Discoverability and Management of Atmospheric Data at Scale: Solutions from the ARM Data Center		Chirag Shah, Wade Darnell, Hannah Collier, Harold Shanafield, Michael Giansiracusa, and Giri Prakash	Short
BigD 436	Beyond Quantile Methods: Improved Top-K Threshold Estimation for Traditional and Learned Sparse Indexes		Jinrui Gou, Yifan Liu, Minghao Shao, and Torsten Suel	Short
BigD 520	KG-CF: Knowledge Graph Completion with Context Filtering under the Guidance of Large Language Models		Zaiyi Zheng, Yushun Dong, Song Wang, Haochen Liu, Qi Wang, and Jundong Li	Short

BigD 604	Multiscale Dubuc: A New Similarity Measure for Time Series	Mahsa Khazaei, Azim Ahmadzadeh, and Krishna Rukmini Puthucode	Short
BigD 298	DifStoryGen: Diffusion-Based Storytelling Algorithm with Distributed Attention	Alireza Nouri and M. Shahriar Hossain	Short
BigD 494	Federated Contrastive Learning of Graph-Level Representations	Xiang Li, Gagan Agrawal, Rajiv Ramnath, and Ruoming Jin	Short
D3S1: Link and Graph Mining II			
BigD 652	Norma: A Framework for Finding Threshold Associations Between Continuous Variables Using Point-wise Functions	Md Mahin and Christoph F Eick	Short
BigD 321	Enhancing Graph Neural Networks with Limited Labeled Data by Actively Distilling Knowledge from Large Language Models	Quan Li, Tianxiang Zhao, Lingwei Chen, Junjie Xu, and Suhang Wang	Short
BigD 681	Feature Interaction Detection in Big Data Through a New Choquet Integral based Deep Neural Network	Matthew Fried, Honggang Wang, and Hua Fang	Short
BigD 682	Explainable Enrichment-Driven GrAph Reasoner (EDGAR) for Large Knowledge Graphs with Applications in Drug Repurposing	Olawumi Olasunkanmi, Evan Morris, Yaphet Kebede, Harlin Lee, Stanley Ahalt, Alexander Tropsha, and Chris Bizon	Short
BigD 747	Evaluating Performance Trade-offs of Caching Strategies for AI-Powered Querying Systems	Hyunju Oh, Wei Zhang, Christopher Rickett, Sreenivas R. Sukumar, and Suren Byna	Short
BigD 882	[Vision Paper] Next Topic Recommendation for Influencers on Social Media	Masafumi Iwanaga, Keishi Tajima, and Yoko Yamakata	Short
BigD 851	WalkBayes: Robust Node Classification via Bayesian Inference and Random Walks under Perturbations	Shuqi He and Jun Song	Short
BigD 873	Efficient counting of balanced (2, k)-bicliques in Signed Bipartite Graphs	Mekala Kiran, Apurba Das, and Suman Banerjee	Short
D1S1: Foundation Models for Big Data			
BigD 497	Benchmarking Human and Automated Prompting in the Segment Anything Model	Jorge Quesada, Zoe Fowler, Mohammad Alotaibi, Mohit Prabhushankar, and Ghassan AlRegib	Regular
BigD 636	Empowering Meta-Analysis: Leveraging Large Language Models for Scientific Synthesis	Jawad Ibn Ahad, Rafeed Mohammad Sultan, Abraham Kaikobad, Fuad Rahman, Mohammad Ruhul Amin, Nabeel Mohammed, and Shafin	Regular

		Rahman	
BigD 646	Prompt Chaining-Assisted Malware Detection: A Hybrid Approach Utilizing Fine-Tuned LLMs and Domain Knowledge-Enriched Cybersecurity Knowledge Graphs	Neha Mohan Kumar, Fahmida Tasnim Lisa, and Sheikh Rabiul Islam	Short
BigD 491	[Vision Paper] Embracing Foundation Models for Advancing Scientific Discovery	Sikun Guo, Amir Shariatmadari, Guangzhi Xiong, and Aidong Zhang	Regular
BigD 216	Automatic Prompt Generation and Optimization by Leveraging Large Language Models to Enhance Few-Shot Learning in Biomedical Tasks	Yiwen Shi and Xiaohua Hu	Regular
BigD 774	Resources Construction and LLM Fine-tuning for Education of Computer Science	Bingjiao Wang, Qinglei Zeng, and Yong Zhang	Regular
D2S1: Foundation Models and Analytics		🥇(Best Paper Award)	
BigD 889	BanglaDialecto: An End-to-End AI-Powered Regional Speech Standardization	Md. Nazmus Sadat Samin, Jawad Ibn Ahad, Tanjila Ahmed Medha, Fuad Rahman, Mohammad Ruhul Amin, Nabeel Mohammed, and Shafin Rahman	Regular
BigD 647	Performance Characterization of Expert Router for Scalable LLM Inference	Josef Pichlmeier, Philipp Ross, and Andre Luckow	Short
BigD 827	Data Augmentations to support Speculative Reasoning in LLMs	Raquib Bin Yousuf, Nicholas Defelice, Mandar Sharma, Shengzhe Xu, and Naren Ramakrishnan	Regular 🥇 (Best Paper Award)
BigD 440	TabularFM: An Open Framework For Tabular Foundational Models	Quan Tran, Suong Hoang, Lam Nguyen, Dzung Phan, and Thanh Lam Hoang	Short
BigD 252	A Study of Foundation Models for Large-scale Time-series Forecasting	Hongjie Chen, Ryan Rossi, Sungchul Kim, Kanak Mahadik, and Hoda Eldardiry	Short
BigD 852	Zero-shot LLM-guided Counterfactual Generation: A Case Study on NLP Model Evaluation	Amrita Bhattacharjee, Raha Moraffah, Joshua Garland, and Huan Liu	Short
BigD 507	Beyond Human Vision: The Role of Large Vision Language Models in Microscope Image Analysis	Prateek Verma, Minh-Hao Van, and Xintao Wu	Short
BigD 816	StyleRec: A Benchmark Dataset for Prompt Recovery in Writing Style Transformation	Shenyang Liu, Yang Gao, Shaoyan Zhai, and Liqiang Wang	Short

BigD 866	Mitigating Sycophancy in Large Language Models via Direct Preference Optimization	Azal Ahmad Khan, Sayan Alam, Xinran Wang, Ahmad Faraz Khan, Debanga Raj Neog, and Ali Anwar	Short
BigD 872	SurgBox: Agent-Driven Operating Room Sandbox with Surgery Copilot	Jinlin Wu, Xusheng Liang, Xuexue Bai, and Zhen Chen	Short
BigD 835	Personalized Federated Learning Techniques: Empirical Analysis	Azal Ahmad Khan, Ahmad Faraz Khan, Haidar Ali, and Ali Anwar	Short
	D1S7: Data Ecosystem		
BigD 818	Calibrating Practical Privacy Risks for Differentially Private Machine Learning	Yuechun Gu and Keke Chen	Regular
BigD 238	Vision Paper: Are Identity Verification Systems Adequately Tailored for Neurodiverse Individuals?	Emanuela Marasco, Nora McDonald, and Vivian Genaro Motti	Short
BigD 330	DP-TabICL: In-Context Learning with Differentially Private Tabular Data	Alycia Noel Carey, Karuna Bhaila, Kennedy Edemacu, and Xintao Wu	Short
BigD 350	SecureSphere: Advancing Security and Robustness in Query Processing over Outsourced Data	Eray Dogan, Meghdad Mirabi, and Carsten Binnig	Short
BigD 736	XSub: Explanation-Driven Adversarial Attack against Black-box Classifiers via Feature Substitution	Kiana Vu, Phung Lai, and Truc Nguyen	Short
BigD 746	Factify: An Automated Fact-Checker for Web Information	Ademola Adesokan and Shady Elbassuoni	Short
BigD 885	Reconsider Time Series Analysis for Insider Threat Detection	Chia-Cheng Chen and Hsing-Kuo Pao	Short
BigD 716	HERL: Tiered Federated Learning with Adaptive Homomorphic Encryption using Reinforcement Learning	Jiaxiang Tang, Zeshan Fayyaz, Mohammad Salahuddin, Raouf Boutaba, Zhi-Li Zhang, and Ali Anwar	Short
BigD 761	Challenging Fairness: A Comprehensive Exploration of Bias in LLM-Based Recommendations	Shahnewaz Karim Sakib and Anindya Bijoy Das	Short
BigD 808	Reliable Time Series Counterfactual Explanations Guided by ShapeDBA	Peiyu Li, Omar Bahri, Soukaina Filali Boubrahimi, and Shah Muhammad Hamdi	Short
BigD 843	Social Biases in Knowledge Representations of Wikidata separates Global North from Global South	Paramita Das, Sai Keerthana Karnam, Aditya Soni, and Animesh Mukherjee	Short
BigD 296	Adversarial Reweighting Guided by Wasserstein Distance to Achieve Demographic Parity	Xuan Zhao, Simone Fabbrizzi, Paula Reyero	Short

		Lobo, Siamak Ghodsi, Klaus Broelemann, Steffen Staab, and Gjergji Kasneci	
BigD 306	Hierarchical Federated Learning with Privacy	Varun Chandrasekaran, Suman Banerjee, Diego Perino, and Nicolas Kourtellis	Regular
BigD 525	RCDP: A Privacy-Preserving Approach for Synthesizing Realistic Commuting Data	Xiaoying Liu, Juanjuan Zhao, Xitong Gao, and Kejiang Ye	Regular
BigD 284	FedGSDW: Enhancing Federated Learning Robustness against Model Poisoning Attack	Xiujin Shi, Naiwen Sun, Jiawei Gong, and Shoujian Yu	Short
	D2S2: Learning from Complex Data		
BigD 766	NysAct: A Scalable Preconditioned Gradient Descent using Nystrom Approximation	Hyunseok Seung, Jaewoo Lee, and Hyunsuk Ko	Short
BigD 768	A Progressive Meta-Algorithm for Large and Seamless Super Resolution Images	Jayalakshmi Mangalagiri, Aryya Gangopadhyay, and David Chapman	Short
BigD 697	Large Scale Evaluation of Deep Learning-based Explainable Solar Flare Forecasting Models with Attribution-based Proximity Analysis	Temitope Adeyeha, Chetraj Pandey, and Berkay Aydin	Short
BigD 775	Impact of Noise in Large Real-World Datasets on Semi-Supervised Object Detection: A Case Study of Homeless Encampments Detection	Bhavyesh Sajja and Seon Ho Kim	Short
BigD 671	Boosting Time Series Prediction of Extreme Events by Reweighting and Fine-tuning	Jimeng Shi, Azam Shirali, and Giri Narasimhan	Short
BigD 777	Spatial-temporal Graph Transformer Network for Spatial-temporal Forecasting	Minh-Son Dao, Koji Zettsu, and Duy-Tang Hoang	Short
BigD 788	Benchmarking in the Service of the End-User	Sarah Alnegheimish, Laure Berti-Équille, and Kalyan Veeramachaneni	Short
BigD 813	Clustering and Topic Discovery of Multiway Data via Joint-NCMTF	Benjamin Cobb, Ricardo Velasquez, Richard Vuduc, and Haesun Park	Short
BigD 826	ACTS: Adaptive Counterfactual Explanations for Time Series Data Using Barycenters [Vision Paper]	Pouya Hosseinzadeh, Soukaina Filali Boubrahimi, and Shah Muhammad Hamdi	Short
	D1S2: Big Data Learning and Analytics I		
BigD 228	Unraveling Intricacies: A Decomposition Approach for Few-Shot Multi-Intent Spoken Language	Wenbin Hua, Yufan Wang, Rui Fan, Xinhui Tu, and	Regular

	Understanding	Tingting He	
BigD 859	Context-Aware Diffusion-based Sequential Recommendation	Di You and Kyumin Lee	Regular
BigD 247	Early Stopping Based on Repeated Significance	Eric Bax, Arundhyoti Sarkar, and Alex Shtoff	Regular
BigD 279	LIVE-ITS: LSH-based Interactive Visualization Explorer for Large-Scale Incomplete Time Series	Hongjie Chen, Aaron Beachnau, Panos Thomas, Pranav Maneriker, Josh Kimball, and Ryan Rossi	Regular
BigD 355	A Green Learning Approach to Efficient Image Demosaicking	Mahtab Movahhedrad, Zijing Chen, and C.-C. Jay Kuo	Regular
BigD 253	Aligning the Representation of Knowledge Graph and Large Language Model for Causal Question Answering	Zefan Zeng, Qing Cheng, Xingchen Hu, Zhong Liu, Jingke Shen, and Yahao Zhang	Regular
BigD 465	Visual Lifelog Retrieval through Captioning-Enhanced Interpretation	Yu-Fei Shih, An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen	Short
D1S8: Big Data Learning and Analytics II			
BigD 369	Automating Data Science Pipelines with Tensor Completion	Shaan Pakala, Bryce Graw, Dawon Ahn, Tam Dinh, Mehnaz Tabassum Mahin, Vassilis Tsotras, Jia Chen, and Evangelos E. Papalexakis	Regular
BigD 415	MAcPNN: Mutual Assisted Learning on Data Streams with Temporal Dependence	Federico Giannini and Emanuele Della Valle	Regular
BigD 438	Improving CLIP for Biomedical Retrieval: A Multi-Margin Contrastive Loss Approach	Ejan Shakya, Haritha Prasad Rayakota, and Pei-Chi Huang	Regular
BigD 288	Unsupervised Machine Learning for Detecting and Locating Human-Made Objects in 3D Point Cloud	Hong Zhao, Huyunting Huang, Tonglin Zhang, Baijian Yang, Jin Wei-Kocsis, and Songlin Fei	Short
BigD 798	Near Real-Time Wildfire Damage Assessment using Aerial Thermal Imagery and Machine Learning	Saqib Azim, Mai Nguyen, Daniel Crawl, Jessica Block, Rawaf Al Rawaf, Francesca Hart, Mark Campbell, Robert Scott, and Ilkay Altintas	Short
BigD 385	Leveraging Spatio-Temporal Locality in Linear Model Trees for Multi-Step Time Series Forecasting	Annunziata D'Aversa, Gianvito Pio, and	Short

		Michelangelo Ceci	
BigD 434	A Comprehensive Forecasting Framework based on Multi-Stage Hierarchical Forecasting Reconciliation and Adjustment	Zhengchao Yang, Mithun Ghosh, Anish Saha, Dong Xu, Konstantin Shmakov, and Kuang-chih Lee	Regular
BigD 387	Fast Second-order Method for Neural Networks under Small Treewidth Setting	Xiaoyu Li, Jiangxuan Long, Zhao Song, and Tianyu Zhou	Regular
D1S9: Big Data Learning and Analytics III			
BigD 448	Accurate Coupled Tensor Factorization with Knowledge Graph	SeungJoo Lee, Yong-chan Park, and U Kang	Regular
BigD 476	A Domain-Agnostic Neurosymbolic Approach for Big Social Data Analysis: Evaluating Mental Health Sentiment on Social Media during COVID-19	Vedant Khandelwal, Manas Gaur, Ugur Kursuncu, Valerie Shalin, and Amit Sheth	Regular
BigD 540	Attention Based Machine Learning Methods for Data Reduction with Guaranteed Error Bounds	Xiao Li, Jaemoon Lee, Anand Rangarajan, and Sanjay Ranka	Regular
BigD 365	Periscope: A Framework for Visualizations of Multiresolution Spatiotemporal Data at Scale	Everett Lewark, Matthew Young, Paahuni Khandelwal, Sangmi Pallickara, and Shrideep Pallickara	Short
BigD 366	LLM4cast: Repurposed LLM for Viral Disease Forecasting	Farah Saeed, Mohammed Aldosari, Budak Arpinar, and John A. Miller	Short
BigD 431	FisherMask: Enhancing Neural Network Labeling Efficiency in Image Classification Using Fisher Information	Shreen Gul, Mohamed Elmahallawy, Sanjay Madria, and Ardhendu Tripathy	Short
BigD 404	COCALITE: A Hybrid Model COmbining CAtch22 and LITE for Time Series Classification	Oumaima Badi, Maxime Devanne, Ali Ismail-Fawaz, Javidan Abdullayev, Vincent Lemaire, Stefano Berretti, Jonathan Weber, and Germain Forestier	Short
BigD 441	Perturbation-driven data augmentation for time series anomaly detection improvement in predictive maintenance	Hyeyoung Lee, Sangkyun Lee, and Sungjoon Choi	Short
BigD 555	VideoBadminton: A Video Dataset for Badminton Action Recognition	Qi Li, Tzu-Chen Chiu, Hsiang-Wei Huang, Min-Te Sun, and Wei-Shinn Ku	Short

BigD 482	Negative Sample Enhancement Strategy to Improve Contrastive Learning for Unsupervised Sentence Representation	Hao Li and Chunzhi Xie	Short
BigD 549	Optimizing Liquid Neural Networks: A comparative study of LTCs and CFCs	Pranav Deepak Tanna, Siddharth Khare, Sanyam Sanjay Jain, Shivodit Raj Vishnoi, Samay Roy, and Jagat Sesh Challa	Short
BigD 413	DataInsight : Big Data Analytics Services	Helia Hedayati and Saeed Samet	Short
BigD 496	Federated Learning under Sample Selection Heterogeneity	Huy Mai and Xintao Wu	Short
BigD 548	Ternary Logit Distillation via Non-Target Classes Decomposition	Ninghua Dong, Zhen Liu, Xiang He, and Yiran Li	Short
BigD 402	SkipSNN: Efficiently Classifying Spike Trains with Event-attention	Hang Yin, Yao Su, Liping Liu, Thomas Hartvigsen, Xin Dai, and Xiangnan Kong	Short
BigD 311	Deep Learning Service for Efficient Data Distribution Aware Sorting	Xiaoke Zhu, Qi Zhang, Wei Zhou, and Ling Liu	Short
BigD 393	CATS: Contrastive learning for Anomaly detection in Time Series	Joël Roman Ky, Bertrand Mathieu, Abdelkader Lahmadi, and Raouf Boutaba	Short
BigD 541	Incomplete Multi-kernel k-means Clustering With Fractional-order Embedding	Deheng Xu, Yun Li, Yun-Hao Yuan, Jipeng Qiang, and Yi Zhu	Regular
	D2S6: Big Data Analytics IV		
BigD 557	A Semi-Supervised Model for Non-Cellular Elements Segmentation in Microscopy Images of Wood	Zihao Xu, Peter Nordström, Sina Sheikholeslami, Ahmad Al-Shishtawy, and Vladimir Vlassov	Short
BigD 870	Risk Diversification Strategy with Moving Average Reversion for Automatic Portfolio Optimization	Yuki Hayashi and Atsuyoshi Nakamura	Regular
BigD 537	GAN Stabilization Under Practical Training Assumptions	Joshua DeOliveira, Walter Gerych, and Elke Rundensteiner	Regular
BigD 578	Edge Grading in Trading Cards Using Transfer Learning: Methods, Experiments, and Evaluation	Lutfun Nahar, Md. Saiful Islam, Mohammad Awrangjeb, and Rob Verhoeve	Short
BigD 596	Global-local Fourier Neural Operator for Accelerating Coronal Magnetic Field Model	Yutao Du, Qin Li, Raghav Gnanasambandam, Mengnan Du, Haimin Wang, and Bo Shen	Short

BigD 731	Semi-Supervised Multi-Source Sea Ice Classification in Small-Data Regime	Samira Alkaee Taleghan, Morteza Karimzadeh, Andrew P.Barrett, Walter N.Meier, and Farnoush Banaei-Kashani	Short
BigD 510	Educational Digital Twin: Tackling Complexity in Educational Big Data	Luwen Huang and Karen Willcox	Short
BigD 456	MIBR: Bridging Domains through Diverse Interests for Cross-Domain Sequential Recommendation	Chengzhe Zhang, Xu Min, Changsheng Li, Xiaolu Zhang, Weichang Wu, Jun Zhou, Ye Yuan, and Guoren Wang	Regular
BigD 572	Meta-Learning for Debiasing Recommendation using Simulated Uniform Data	sibo lu, Zhen Liu, xinxin yang, yilin ding, yibo gao, and yafan yuan	Short
BigD 650	Divide2Conquer (D2C): A Decentralized Approach Towards Overfitting Remediation in Deep Learning	Md. Saiful Bari Siddiqui, Md Mohaiminul Islam, and Md. Golam Rabiul Alam	Short
BigD 584	Alternative Methods to SHAP Derived from Properties of Kernels: A Note on Theoretical Analysis	Kazuhiro Hiraki, Shinichi Ishihara, and Junnosuke Shino	Short
BigD 556	Improving Multi-Domain Task-Oriented Dialogue System with Offline Reinforcement Learning	Dharmendra Prajapat and Durga Toshniwal	Short
BigD 616	A Data-Driven Approach for Automated Multi-Site Competitive Facility Location	Ming Hui Tan, Kar Way Tan, and Hoong Chuin Lau	Short
BigD 713	Commercial Space Classification and Prediction with Machine Learning and Multiple Data Sources: A Case Study of Milan	Jiabin Wei	Short
BigD 724	DANCE: Deep Learning-Assisted Analysis of ProteiN Sequences Using Chaos Enhanced Kaleidoscopic Images	Taslim Murad, Prakash Chourasia, Sarwan Ali, and Murray Patterson	Short
	D1S10: Deep Learning I		
BigD 558	[Vision Paper] Quality-Aware Experience Exploitation in Model-Based Reinforcement Learning	Guang Yang, Jiahe Li, Ziye Geng, and Changqing Luo	Regular
BigD 559	A Feedback-based Decision-Making Mechanism for Actor-Critic Deep Reinforcement Learning	Guang Yang, Ziye Geng, Jiahe Li, Yanxiao Zhao, Sherif Abdelwahed, and Changqing Luo	Regular
BigD 580	Explingo: Explaining AI Predictions using Large Language Models	Alexandra Zytek, Sara Pido, Sarah Alnegheimish, Laure Berti-Equille, and Kalyan Veeramachaneni	Regular
BigD 588	Data-driven Synchronization Protocols for Data-parallel Neural Learning over Streaming Data	George Klioumis and Nikos Giatrakos	Regular

BigD 601	SeriesGAN: Time Series Generation via Adversarial and Autoregressive Learning	MohammadReza EskandariNasab, Shah Muhammad Hamdi, and Soukaina Filali Boubrahimi	Regular
	D1S11: Machine Learning Algorithms I		
BigD 623	Stochastic Scale Invariant Power Iteration for KL-divergence Nonnegative Matrix Factorization	Cheolmin Kim, Youngseok Kim, Yegna Subramanian Jambunath, and Diego Klabjan	Regular
BigD 633	UnifiedGT: Towards a Universal Framework of Transformers in Large-Scale Graph Learning	Junhong Lin, Xiaojie Guo, Shuaicheng Zhang, Dawei Zhou, Yada Zhu, and Julian Shun	Regular
BigD 641	OSR-ViT: A Simple and Modular Framework for Open-Set Object Detection and Discovery	Matthew Inkawhich, Nathan Inkawhich, Hao Yang, Jingyang Zhang, Randolph Linderman, and Yiran Chen	Regular
BigD 648	ORIS: Online Active Learning Using Reinforcement Learning-based Inclusive Sampling for Robust Streaming Analytics System	Rahul Pandey, Ziwei Zhu, and Hemant Purohit	Regular
BigD 663	Enhancing Big Data Analysis: A Recursive Window Segmentation Strategy for Multivariate Longitudinal Data	Desmond Fomo and Aki-Hiro Sato	Regular
	D1S12: Deep learning II	🏅Runner-up for Best Paper Award	
BigD 668	Targeting Negative Flips in Active Learning using Validation Sets	Ryan Benkert, Mohit Prabhushankar, and Ghassan AlRegib	Regular 🏅Runner-up for Best Paper Award
BigD 684	PlugVFL: Robust and IP-Protecting Vertical Federated Learning against Unexpected Quitting of Parties	Jingwei Sun, Zhixu Du, Anna Dai, Saleh Baghersalimi, Alireza Amirshahi, David Atienza, and Yiran Chen	Regular
BigD 719	Effective Guidance for Model Attention with Simple Yes-no Annotations	Seongmin Lee, Ali Payani, and Duen Horng Chau	Regular
BigD 720	SPrint: Self-Paced Continual Learning with Adaptive Curriculum and Memory Replay	Min-Seon Kim, Ling Liu, and Hyuk-Yoon Kwon	Regular
BigD 725	LPLgrad: Optimizing Active Learning Through Gradient Norm Sample Selection and Auxiliary Model Training	Shreen Gul, Mohamed Elmahallawy, Sanjay Madria, and Ardhendu Tripathy	Regular
	D2S10: Machine Learning	🏅Best Student Paper	

	Algorithms II	Award	
BigD 753	Predicting Surface Water Bacteria Levels Using Transfer Learning and Domain Adaptation	Ali Elahi, David Shumway, Megan Kowalcyk, Abhilasha Shrestha, Nikita Gautam, Doina Caragea, Cornelia Caragea, and Samuel Dorevitch	Regular
BigD 791	Efficient Hierarchical Contrastive Self-supervising Learning for Time Series Classification via Importance-aware Resolution Selection	Kevin Garcia, Juan Perez, and Yifeng Gao	Regular
BigD 796	SplitVAEs: Decentralized scenario generation from siloed data for stochastic optimization problems	H M Mohaimanul Islam, Huynh Quang Nguyen Vo, and Paritosh Ramanan	Regular
BigD 319	OL4TeX: Adaptive Online Learning for Text Classification under Distribution Shifts	Min-Seon Kim, Ling Liu, and Hyuk-Yoon Kwon	Short
BigD 250	Fine-grained Graph-based Anomaly Detection on Vehicle Controller Area Networks	Isaiah J. King, Benjamin Bowman, and H. Howie Huang	Short
BigD 628	RYAN: A tool for explaining and visually analyzing the evolution of Relaxed Functional Dependencies	Loredana Caruccio, Stefano Cirillo, Gianpaolo Iuliano, Giuseppe Polese, and Roberto Stanzione	Short
BigD 764	A Hybrid Model Based on Graph Convolutional Network and Multi-head Transformer for Joint Multiple Intent Detection and Slot Filling	Dongsheng Li, Kun Niu, Xiao Chen, and Zhihai Wang	Regular
BigD 733	DynamicFL: Federated Learning with Dynamic Communication Resource Allocation	Qi Le, Enmao Diao, Xinran Wang, Ahmad Faraz Khan, Vahid Tarokh, Jie Ding, and Ali Anwar	Regular 🥇 Best Student Paper Award
	D3S2: Machine Learning Algorithms III	🎗 Runner-up for Best Student PAper Award	
BigD 801	Pre-Trained Language Models with Topic Attention for Supervised Document Structure Learning	Dang Pham and Tuan Le	Regular
BigD 858	MADOD: Generalizing OOD Detection to Unseen Domains via G-Invariance Meta-Learning	Haoliang Wang, Chen Zhao, and Feng Chen	Regular
BigD 832	Real-Time Video Analysis for Accurate Biometric Identity Verification in Mobile Devices	Sulabh Shrestha, Emanuela Marasco, and Babek Norouzlou	Regular
BigD 840	Tenet: Benchmarking Data Stream Classifiers in Presence of Temporal Dependence	Giacomo Ziffer, Federico Giannini, and Emanuele Della Valle	Regular 🎗 Runner-up for Best Student PAper Award
BigD 856	FEED: Fairness-Enhanced Meta-Learning for Domain Generalization	Kai Jiang, Chen Zhao, Haoliang Wang, and Feng Chen	Regular

BigD 817	2FWL-SIRGN: A Scalable Structural 2-dimensional Folklore Weisfeiler Lehman Graph Representation Learning Approach Via Structural Graph Partitioning	Justin Carpenter and Edoardo Serra	Regular
BigD 251	Personalized Federated Learning by Domain-Aware Network Pruning and Re-growth	Yuto Suzuki and Farnoush Banaei-Kashani	Short
BigD 763	SEFD: Semantic-Enhanced Framework for Detecting LLM-Generated Text	Weiqing He, Bojian Hou, Tianqi Shang, Davoud Ataee Tarzanagh, Qi Long, and Li Shen	Short
BigD 689	Data Reliability Enhanced Prediction for Recommendation System: A Case Study on Named Entity Recognition	Prianka Banik, Lin Li, Xishuang Dong, and Lijun Qian	Short
BigD 702	Deep Learning for Koopman Operator Estimation in Idealized Atmospheric Dynamics	David Millard, Arielle Carr, and Stephane Gaudreault	Short
BigD 706	Efficient SGD Neural Network Training via Sublinear Activated Neuron Identification	Lianke Qin, Zhao Song, and Yuanyuan Yang	Short
BigD 219	Initialization Method for t-Distributed Stochastic Neighbour Embedding (t-SNE)	Sara Nassar, Samir Belhaouari, Mebarka Allaoui, and Mohammed Kherfi	Short
	D1S13: Big Data Applications I		
BigD 227	Predicting ChatGPTs Ability to Solve Complex Programming Challenges	Nguyen Ho, James May, Bao Ngo, Jack Formato, Linh Ngo, Van Long Ho, and Hoang Bui	Regular
BigD 323	A Confidence-Based Power-Efficient Framework for Sleep Stage Classification on Consumer Wearables	Hsu-Chao Lai, Po-Hsiang Fang, Yi-Ting Wu, Lo Pang-Yun Ting, and Kun-Ta Chuang	Regular
BigD 352	Evaluate the Policy Impact of Temporary Suspension of Social Media Account using Synthetic Control Method	King wa Fu	Regular
BigD 331	SparrowVQE: Visual Question Explanation for Course Content Understanding	Jialu Li, Manish Thota, Ruslan Gokhman, Radek Holik, and Youshan Zhang	Regular
BigD 340	Is it Fine to Tune? Evaluating SentenceBERT Fine-tuning for Brazilian Portuguese Text Stream Classification	Bruno Yuiti Leão Imai, Cristiano Mesquita Garcia, Marcio Vinicius Rocha, Alessandro Lameiras Koerich, Alceu de Souza Britto Jr., and Jean Paul Barddal	Regular
BigD 334	Enhancing Financial Reasoning in Large Language Models: The Role of Gold Facts	Shoutai Zhu, Ziqiang Yuan, Kaiyuan Wang, Yishu Zhang, and Wenqi	Regular

		Wei	
	D1S14: Big Data Applications II		
BigD 348	EEG2TEXT: Open Vocabulary EEG-to-Text Translation with Multi-View Transformer	Hanwen Liu, Daniel Hajialigol, Benny Antony, Aiguo Han, and Xuan Wang	Regular
BigD 354	Dynamic User Grouping and Evolution Tracking (DUGET): Leveraging Machine Learning for Public Transit Insights	Tobias Johannesson, Isak Rubensson, Sina Sheikholeslami, Ahmad Al-Shishtawy, and Vladimir Vlassov	Regular
BigD 408	[Vision Paper] PRObot: Enhancing Patient-Reported Outcome Measures for Diabetic Retinopathy using Chatbots and Generative AI	Maren Pielka, Tobias Schneider, Jan Terheyden, and Rafet Sifa	Regular
BigD 468	Deep Computer Vision for Solar Physics Big Data: Opportunities and Challenges [Vision Paper]	Bo Shen, Marco Marena, Chenyang Li, Qin Li, Haodi Jiang, Mengnan Du, Jiajun Xu, and Haimin Wang	Regular
BigD 606	A Scalable Game-theoretic Approach to Urban Evacuation Routing and Scheduling	Kazi Ashik Islam, Da Qi Chen, Madhav Marathe, Henning Mortveit, Samarth Swarup, and Anil Vullikanti	Regular
BigD 450	Accurate Stock Movement Prediction via Multi-Scale and Multi-Domain Modeling	JinGee Kim, Yong-chan Park, Jaemin Hong, and U Kang	Regular
BigD 626	Melanoma Classification using GAN based augmentation and Self-Supervised feature extraction	Akanksha Lal, Sadhana Tiwari, Rushil Patra, and Sonali Agrawal	Short
BigD 615	EGANKT: Enhancing Graph-Attention Networks for Knowledge Tracing by Predicting Concepts and Abilities	Qiwen Zheng, Shun Mao, Kai Chen, and Yuncheng Jiang	Regular
BigD 388	Constructing A Human-Centered Dataset for Second Language Acquisition via Large Language Models	Yunlong Fan, Baixuan Li, Zhiheng Yang, and Zhiqiang Gao	Regular
BigD 639	DistDNAS: Search Efficient Feature Interactions within 2 Hours	Tunhou Zhang, Wei Wen, Igor Fedorov, Xi Liu, Buyun Zhang, Fangqiu Han, Wen-Yen Chen, Yiping Han, Feng Yan, Hai Li, and Yiran Chen	Short
	D1S15: Big Data Analytics I	🎗Honorable Mention for Best Paper Award	
BigD 659	LongKey: Keyphrase Extraction for Long Documents	Jeovane Honorio Alves, Radu State, Cinthia Obladen de Almendra	Regular

		Freitas, and Jean Paul Barddal	
BigD 698	Animal-JEPA: Advancing Animal Behavior Studies Through Joint Embedding Predictive Architecture in Video Analysis	Chengjie Zheng, Tewodros Mulugeta Dagnew, Liuyue Yang, Wei Ding, Shiqian Shen, Changning Wang, and Ping Chen	Regular
BigD 597	susFL: Federated Learning-based Monitoring for Sustainable, Attack-Resistant Smart Farms	Dian Chen, Paul Yang, Dong Sam Ha, and Jin-Hee Cho	Regular
BigD 749	An Active Learning-Based Streaming Pipeline for Reduced Data Training of Structure Finding Models in Neutron Diffractometry	Tianle Wang, Jorge Ramirez, Cristina Garcia-Cardona, Thomas Proffen, Shantenu Jha, and Sudip Seal	Regular 🎗Honorable Mention for Best Paper Award
BigD 769	Enhancing Cross-Modal Contextual Congruence for Crowdfunding Success using Knowledge Graphs	Trilok Padhi, Ugur Kursuncu, Yaman Kumar, Valerie L. Shalin, and Lane Peterson Fronczek	Regular
BigD 696	SE(3) Equivariant Neural Network for 3D Graphs	Sarp Aykent and Tian Xia	Regular
D1S16: Big Data Analytics II			
BigD 799	Efficient Arrhythmia Detection Using Progressive Resolution Shrinking	Tavonput Luangphasy and Xinghui Zhao	Regular
BigD 838	Context-Specific Feature Augmentation for Improving Social Determinants of Health Extraction	Lei Gong, Aidong Zhang, Andrey Shor, and Kishlay Jha	Regular
BigD 687	PerSSD: Persistent, Shared, and Scalable Data with Node-Local Storage for Scientific Workflows in Cloud Infrastructure	Paula Olaya, Sophia Wen, Jay Lofstead, and Michela Taufer	Regular
BigD 846	Understanding Student Sentiment on Mental Health Support in Colleges Using Large Language Models	Palak Sood, Chengyang He, Divyanshu Gupta, Yue Ning, and Ping Wang	Regular
BigD 834	SO(3) Equivariant Framework for Spatial Networks	Tian Xia and Sarp Aykent	Regular
D1S17: Big Data Analytics III			
BigD 406	Reject Inference as a Noisy Label Detection and Counterfactual Correction Task	Charalampos Chelmis, Mahsa Azarshab, Khandker Sadia Rahman, and Mehrdad Mirpourian	Short
BigD 407	Multimodal survival prediction using TabTransformer and BioClinicalBERT on MIMIC-III	Beatrice Insalata, Fabian Schmidt, and Vladimir Vlassov	Short

BigD 477	T2I-RISE - A Comprehensive Approach with Conversational Context and Feedback Systems for Text-to-Insights with Reinforcement learning, Integration of Semantic layers and Enrichment	Raghu Chandra Katikeri, Sai Phaniraja Maddila, Sai Pratheek, Rajvi Desai Birenbhai, Amit Vaid, Neelesh Shukla, and Sandeep Jain	Short
BigD 829	Multimodal Learning on Temporal Data	Ye Xue, Diego Klabjan, and Jean Utke	Short
BigD 755	Estimate Causal Effects of Entangled Treatment on Graphs using Disentangled Instrumental Variables	JINGYUAN CHOU, JIANGZHUO CHEN, and MADHAV MARATHE	Short
BigD 760	On Rank Selection for Nonnegative Matrix Factorization	Srinivas Eswar, Koby Hayashi, Benjamin Cobb, Ramakrishnan Kannan, Grey Ballard, Richard Vuduc, and Haesun Park	Short
BigD 479	EVOLVE: Evaluation of Language-to-SQL Validity and Effectiveness - A Detailed Review Framework for Complex Text-to-SQL Queries	Raghu Chandra Katikeri, Alok Nook Raj Pepakayala, Lokesh Kuncham, Rachita Barla, Tarun Kasula, Sai Charan Ravulapally, and Amit Vaid	Short
BigD 503	Noise2Inverse for 3D Low-Dose Cone-Beam Computed Tomography	Austin Yunker, Jason Luce, John C. Roeske, Rajkumar Kettimuthu, Hyejoo Kang, Sebastien Gros, and Alec M. Block	Short
BigD 223	Prioritizing Potential Wetland Areas via Region-to-Region Knowledge Transfer through Domain Disentanglement and Adaptive Propagation	Yoonhyuk Choi, Reepal Shah, John Sabo, Huan Liu, and Selçuk Candan	Short
BigD 274	Construction and Application of the SMART Model for Adaptive Industrial Data Collection Based on Knowledge Graphs	Wendan Cheng, Zhen Zhang, Heng Qian, Qiuyue Wang, Guanqun Su, and Lingge Meng	Short
BigD 564	EXCON: Extreme Instance-based Contrastive Representation Learning of Severely Imbalanced Multivariate Time Series for Solar Flare Prediction Tasks	Onur Vural, Shah Muhammad Hamdi, and Soukaina Filali Boubrahimi	Short
BigD 501	YOLOv10 Computer Vision Performance Measurement for Agricultural Vacuum Seed Meters	Manuel Blaser, Wesley Porter, Luke Fuhrer, and Thirimachos Bourlai	Short
BigD 567	A Triple Factorization-Based Coupled Latent Factor Analysis Model for Recommender Systems	Yunchun Xian, Qicong Hu, and Hao Wu	Short
BigD 327	Data Science and Massive Open Online Courses (MOOCs): Key Feats and Best Practices	Damian Andrew Tamburri and Willem-Jan van den Heuvel	Short

Industry and Government Program Sessions

I&G-1: Generative AI Dec. 16, Monday, 10:30-12:30, Columbia C	
Paper 2 (N221) LLM Chain Ensembles for Scalable and Accurate Data Annotation (regular)	David Farr, Nico Manzonelli, Iain Cruickshank, Kate Starbird, and Jevin West
Paper 6 (N257) Experimental Study of In-Context Learning for Text Classification and Its Application to Legal Document Review in Construction Delay Disputes (regular)	Nathaniel Huber-Fliflet, Jianping Zhang, Peter Gronvall, and Fusheng Wei
Paper 9 (N217) IDNet: A Novel Identity Document Dataset via Few-Shot and Quality-Driven Synthetic Data Generation (regular)	Lulu Xie, Yancheng Wang, Hong Guan, Soham Nag, Rajeev Goel, Niranjan Swamy, Yingzhen Yang, Chaowei Xiao, Jonathan Prisby, Ross Maciejewski, and Jia Zou
Paper 23 (N233) AUTORED: Automated Attack Scenario Generation Framework for Red Teaming of LLMs (short)	Zhe Wang and Mohammad Tayebi
Paper 26 (N253) Generating Phishing Attacks and Novel Detection Algorithms in the Era of Large Language Models (short)	Jeffrey Fairbanks and Edoardo Serra
Paper 28 (N237) Enriching Skill Taxonomies through Vector Space Models (short)	Simone D'Amico, Alessia De Santo, Fabio Mercorio, and Mario Mezzanzanica
I&G-2: Big data Infra & Multi-modality Dec. 16, Monday, 14:00-16:00, Columbia C	
Paper 1 (N239) DIVERSE: A Dataset of YouTube Video Comment Stances with a Data Programming Model (regular)	Iain Cruickshank, Lynnette Ng, and Amir Soofi
Paper 15 (N209) INFA-FinOps for Cloud Data Integration (regular)	Shivangi Srivastava, Anant Mittal, Atam Agrawal, Valentin Moskovich, Michael Brevard, and Mosharaf Chowdhury

Paper 20 (N218) Capturing Infrastructure Interdependencies for Power Outages Prediction During Extreme Events (regular)	Sangkeun Lee, Avishek Bose, Narayan Bhusal, and Supriya Chinthavali
Paper 29 (N204) GPURank: A Cloud GPU Instance Recommendation System (short)	Shravika Mittal, Kanak Mahadik, Ryan Rossi, Sungchul Kim, and Handong Zhao
Paper 32 (N271) Addressing Spend Leakage and Optimization of Cloud Costs (short)	Uday Chandra Bhookya, Kunal Jethuri, Sanjeeva Rayudu Ravuru, Priyadarshi Priyadarshi, and Maitreya Natu
Paper 33 (N262) Forecasting Application Counts in Talent Acquisition Platforms: Harnessing Multimodal Signals using LMs (short)	Makini Chisolm-Straker and Lav R. Varshney
I&G-3: Modeling **Dec. 16, Monday, 14:00-16:00, Concord**	
Paper 5 (N278) Data-Prep-Kit: getting your data ready for LLM application development (regular)	David Wood, Boris Lublinsky, Alexy Roytman, Shivdeep Singh, Constantin Adam, Abdulhamid Adebayo, Sungeun An, Yuan Chi Chang, Xuan-Hong Dang, Nirmit Desai, Michele Dolfi, Hajar Emami-Gohari, Revital Eres, Takuya Goto, Dhiraj Joshi, Yan Koyfman, Mohammad Nassar, Hima Patel, Parameswaran Selvam, Yousaf Shah, Saptha Surendran, Daiki Tsuzuku, Petros Zerfos, and Shahrokh Daijavad
Paper 10 (N272) WeightedMMoE: Weighted Multi-Gate Mixture-of-Experts for Multi-Source Unsupervised Domain Adaptation (regular)	Siti Mariyah and Wayne Wobcke
Paper 12 (N220) HLAT: High-quality Large Language Model Pre-trained on AWS Trainium (regular)	Haozheng Fan, Hao Zhou, Guangtai Huang, Parameswaran Raman, Xinwei Fu, Gaurav Gupta, Dhananjay Ram, Yida Wang, and Jun Huan
Paper 14 (N206) Active Learning for Practical Misinformation Classification in Social Media: a Case Study on COVID-19 (regular)	Han Kyul Kim and Andy Skumanich

Paper 18 (N245) Machine Learning (ML) Classifier to Assist in Metadata Creation (regular)	Hannah Collier, Eric Enright, Sujata Goswami, Chirag Shah, Maggie Davis, and Rachael Isphording
Paper 34 (N236) An explainable framework based on counterfactual explanations for multi-class financial distress prediction of small and medium enterprises) (short)	Renon Ando, Yuji Kawamata, Toshihiko Takeda, and Yukihiko Okada
I&G-4: Business and Energy **Dec. 18, Tuesday, 10:30-12:30, Columbia C**	
Paper 3 (N235) Targeted Detection of Anomalous Merchants on Integrated Payment Platforms via Multifaceted Transaction Representation Learning (regular)	Guanyu Lu, Xiang Lin, Martin Pavlovski, Xinyu Zhang, and Fang Zhou
Paper 4 (N203) End-to-End Trip Energy Predictions Developed from Real-World Driving Data using Recurrent Neural Networks (regular)	Shinhoon Kim, Vishnu Raghupathy, Yongkang Liu, Keisuke Niimi, and Takahiro Mochihara
Paper 7 (N254) Counter Data Paucity through Adversarial Invariance Encoding: A Case Study on Modeling Battery Thermal Runaway (regular)	Anika Tabassum, Srikanth Allu, and Ramakrishnan Kannan
Paper 17 (N244) ProbSAINT: Probabilistic Tabular Regression for Used Car Pricing (regular)	Kiran Madhusudhanan, Gunnar Behrens, Maximilian Stubbemann, and Lars Schmidt-Thieme
Paper 30 (N266) Forecasting Migration Patterns and Land Border Encounters (short)	Raquib Bin Yousuf, Shengzhe Xu, Patrick Butler, Brian Mayer, Nathan Self, David Mares, and Naren Ramakrishnan
I&G-5: Recommender systems **Dec. 18, Wednesday, 10:30-12:30, Columbia A**	
Paper 8 (N256) Multi-task Recommendation in Marketplace via Knowledge Attentive Graph Convolutional Network with Adaptive Contrastive Learning (regular)	Xiaohan Li, Zezhong Fan, Luyi Ma, Kaushiki Nag, and Kannan Achan
Paper 16 (N216) Optimal Transport for Efficient, Unsupervised Anomaly Detection on Industrial Data (regular)	Abigail Langbridge, Fearghal O'Donncha, James Rayfield, and Bradley Eck

Paper 21 (N213) GRAINRec: Graph and Attention Integrated Approach for Real-Time Session-Based Item Recommendations (regular)	Bhavtosh Rath, Pushkar Chennu, David Relyea, Prathyusha Kanmanth Reddy, and Amit Pande
Paper 24 (N224) Exploring Query Understanding for Amazon Product Search (short)	Chen Luo, Xianfeng Tang, Hanqing Lu, Yaochen Xie, Hui Liu, Zhenwei Dai, LImeng Cui, Ashutosh Joshi, Sreyashi Nag, Yang Li, Zhen Li, Rahul Goutam, Jiliang Tang, Haiyang Zhang, and Qi He
Paper 27 (N259) CryptoPulse: Short-Term Cryptocurrency Forecasting with Dual-Prediction and Cross-Correlated Market Indicators (short)	Amit Kumar and Taoran JI
Paper 35 (N229) Harnessing the Power of Graph Neural Networks for Personalized Rail Recommendations (short)	Bora Edizel, Sri Haindavi Koppuravuri, Mark Gannaway, Puja Das, and Kamilia Ahmadi
I&G-6: AI4Science **Dec. 18, Wednesday, 10:30-12:30, Columbia B**	
Paper 11 (N252) Enabling Scalable Mineral Exploration: Self-Supervision and Explainability (regular)	Angel Daruna, Vasily Zadorozhnyy, Georgina Lukoczki, and Han-Pang Chiu
Paper 13 (N265) Accurate and Interpretable Radar Quantitative Precipitation Estimation with Symbolic Regression (regular)	Olivia Zhang, Brianna Grissom, Julian Pulido, Kenia Munoz-Ordaz, Jonathan He, Mostafa Cham, Haotong Jing, Weikang Qian, Yixin Wen, and Jianwu Wang
Paper 19 (N243) Monitoring Recombinant Protein Titer in Escherichia Coli Fermentations Using Filtering Techniques (regular)	Paul Okafor, Zahed Siddique, and Talayeh Razzaghi
Paper 22 (N268) Detecting Fraud in a Large Anonymized Voter Registration Dataset (short)	Nahid Anwar, Amit Jain, Edoardo Serra, and Chad Houck
Paper 25 (N215) ARIM-mdx Data System: Towards a Nationwide Data Platform for Materials Science (short)	Masatoshi Hanai, Ryo Ishikawa, Mitsuaki Kawamura, Masato Ohnishi, Norio Takenaka, Kou Nakamura, Daiju Matsumura, Seiji Fujikawa, Hiroki Sakamoto, Yukinori Ochiai, Tetsuo Okane, Shin-Ichiro Kuroki,

	Atsuo Yamada, Toyotaro Suzumura, Junichiro Shiomi, Kenjiro Taura, Yoshio Mita, Naoya Shibata, and Yuichi Ikuhara
Paper 31 (N202) Data Workbench For Earth and Atmospheric Science Research Community (short)	Sujata Goswami, Kyle Dumas, Wade Darnell, Vasile Tudor Garbulet, Michael Giansiracus, and Giri Prakash
Paper 32 (N241) *Content-Aware Deep Learning Recommender System*	Qixin Wang, Xintao Wu, Kan Yao, and Han Li

Special Sessions

Dec. 15, Sunday			
9:00-17:00	Congressional C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
9:00-17:00	Regency C	Special Session: Federated Learning on Big Data	Francesco Piccialli
9:00-17:00	Sequoia	Special Session: Intelligent Data Mining	Uraz Yavanoglu
Dec. 16, Monday			
8:30-18:00	Bryce	Joint Session Workshop #14: 2nd WS on Large-scale Data Utilization in Economics of Information and Management Sciences: Theory, Computation, and Experiment Special Session: Understanding New Markets by Data Science, Social Science, and Economics https://tetsuwaka.net/UNMDSSSE2024/schedule.html	Teruaki Hayashi Naoki Watanabe Hiroki Sakaji
10:30-18:00	Glacier	Special Session: Federated Learning on Big Data	Francesco Piccialli
10:30-18:00	Congressional C	Special Session: Machine Learning on Big Data	Alfredo Cuzzocrea
10:30-18:00	Sequoia	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea
Dec. 17, Tuesday			
10:30-16:30	Yosemite	Special Session: Synergizing Mobility Data for Creating and Discovering Valuable Places https://www.panda.sys.t.u-tokyo.ac.jp/Syn_Mobil.html	Yukio Ohsawa
10:30-16:30	Grand Teton	Special Session: Social Cognitive Computing in Digital Education and Learning	Jerry Chun-Wei Lin
10:30-16:30	Glacier	Special session: Privacy and Security of Big Data	Alfredo Cuzzocrea

Dec. 18, Wednesday			
10:30-16:00	Congressional A	Special Session: HealthCare Data	Ozgun Pinarer
10:30-12:30	Congressional B	Special Session: Dataspaces and DFFT (Data Free Flow with Trust)	Noboru Koshizuka
10:30-12:30	Yosemite	Special Session: Synergizing Mobility Data for Creating and Discovering Valuable Places	Yukio Ohsawa

PhD Forum

PhD Forum: Poster Session Dec. 17, Tuesday, 10:00-10:30, Regency Foyer Columbia Foyer	
PhD Forum: Paper Session I **Dec. 17, Tuesday, 10:30-11:30, Concord**	
SP13207: Towards Trustworthy Graph Neural Networks and Their Applications in Recommender Systems	Longfeng Wu
SP13225: Towards a Supporting Framework for Neuro-Developmental Disorder: Considering Artificial Intelligence, Serious Games and Eye Tracking	Abdul Rehman, Ilona Heldal, Diana Stilwell, and Jerry Chun-Wei Lin
SP13217: Time Series Causal Discovery Using a Hybrid Method	Saima Absar and Lu Zhang
PhD Forum: Panel **Dec. 17, Tuesday, 11:30-12:30, Concord**	
PhD Forum: Paper Session II **Dec. 17, Tuesday, 14:00-16:00, Concord**	
SP13221: TIFG: Text-Informed Feature Generation with Large Language Models	Xinhao Zhang and Kunpeng Liu
SP13201: Knowledge Transfer Predictive Models for Power Outage Caused by Various Types of Extreme Weather Events	Jangjae Lee and Stephanie Paal
SP13213: Domain-Aware LLM Routing During Generation	Josef Pichlmeier, Philipp Ross, and Andre Luckow
SP13222: Thought Space Explorer: Navigating and Expanding Thought Space for Large Language Model Reasoning	Jinghan Zhang and Kunpeng Liu
SP13208: Feature-Space Semantic Invariance: Enhanced OOD Detection for Open-Set Domain Generalization	Haoliang Wang, Chen Zhao, and Feng Chen
SP13220: User Privacy in Skeleton-based Motion Data	Thomas Carr and Depeng Xu
SP13206: BadSAD: Clean-Label Backdoor Attacks against Deep Semi-Supervised Anomaly Detection	He Cheng

SP13223: Global and Local Structure Learning for Sparse Tensor Completion	Dawon Ahn and Evangelos Papalexakis

<table>
<tr><td colspan="2" align="center">PhD Forum: Poster Session
Dec. 17, Tuesday, 16:00-16:30, Regency Foyer Columbia Foyer</td></tr>
<tr><td colspan="2" align="center">PhD Forum: Paper Session II
Dec. 17, Tuesday, 16:30-17:30, Concord</td></tr>
<tr><td>SP13205: Leveraging Big Data Technologies for Practical Radio Frequency Fingerprinting Applications</td><td>Stefan Kunze and Wolfgang Dorner</td></tr>
<tr><td>SP13211: Optimizing Deployment of Homomorphic Encryption and SQL using Reinforcement Learning</td><td>Ryan Marinelli, Avald Sommervoll, and Laszlo Erdodi</td></tr>
<tr><td>SP13216: Robust Hate Speech Detection Without Predefined Spurious Words</td><td>Xingyi Zhao</td></tr>
<tr><td>SP13212: Accounting for Cancer Patients with Severe Outcomes: An Anomaly Detection Perspective</td><td>Yang Yan, Christopher Lominska, Gregory Gan, Hao Gao, and Zhong Chen</td></tr>
<tr><td>SP13209: Enhancing Customer Behavior Prediction and Interpretability</td><td>Yu-Chung Wang, Lars Arne Jordanger, and Jerry Chun-Wei Lin</td></tr>
</table>

Online Workshop Sessions

Dec. 15, Sunday		
Workshop #11: The 7th International Workshop on "Big Data Engineering and Analytics in Cyber-Physical Systems" (BigEACPS'24) https://sites.google.com/view/bigeacps24	Dec 15, Morning	Chair: Akbar Namin
Workshop #12: The 2nd International Workshop on Exploring Cutting-edge Data Mining: Applications of Infrared Target Detection and Large Language Models https://aidatalab.github.io/bigdata2024/pastworkshopd.html	Dec 15, Morning	Chair: Jinli Zhang
Combined Workshops: #19: Foundation Models for Science Big Data & #37: IEEE 5th Workshop on High Performance Computing, Big Data Analytics, and Integration for Multi-Omics Biomedical Data (HPC-BOD) https://xuanwang91.github.io/fm4science-workshop/ https://hpcbod.cs.fiu.edu/	Dec 15, Morning	Chairs: Xuan Wang & Fahad Saeed
Workshop #25: IEEE Workshop on Big Data Analytics for Medical Imaging https://sites.google.com/unisa.it/bdami2024	Dec 15, Morning	Chair: Carmen Bisogni
Workshop #35: The IEEE International Workshop on Large Language Models for Finance https://intelligentfinance.github.io/IEEE-LLM-finance-2024/index.html	Dec 15, Whole Day	Chair: Quanzhi Li
Workshop #40: Workshop on Handling Resource Constraints for/using Bigdata and AI https://lc4workshop.github.io/constraint.html	Dec 15, Morning	Chair: Manikandan Ravikiran
Workshop #45: 2nd International Workshop on Dataspaces and Digital Twins for Critical Entities and Smart Urban Communities (DSpaCES) https://sites.google.com/unisalento.it/dspaces2024	Dec 15, Morning	Chair: Marco Zappatore
Workshop #49: The 8th IEEE International Workshop on Big Spatial Data (BSD 2024) https://confeera.com/conf/4/bsd-2024/about	Dec 15, Morning	Chair: Ashwin Shashidharan

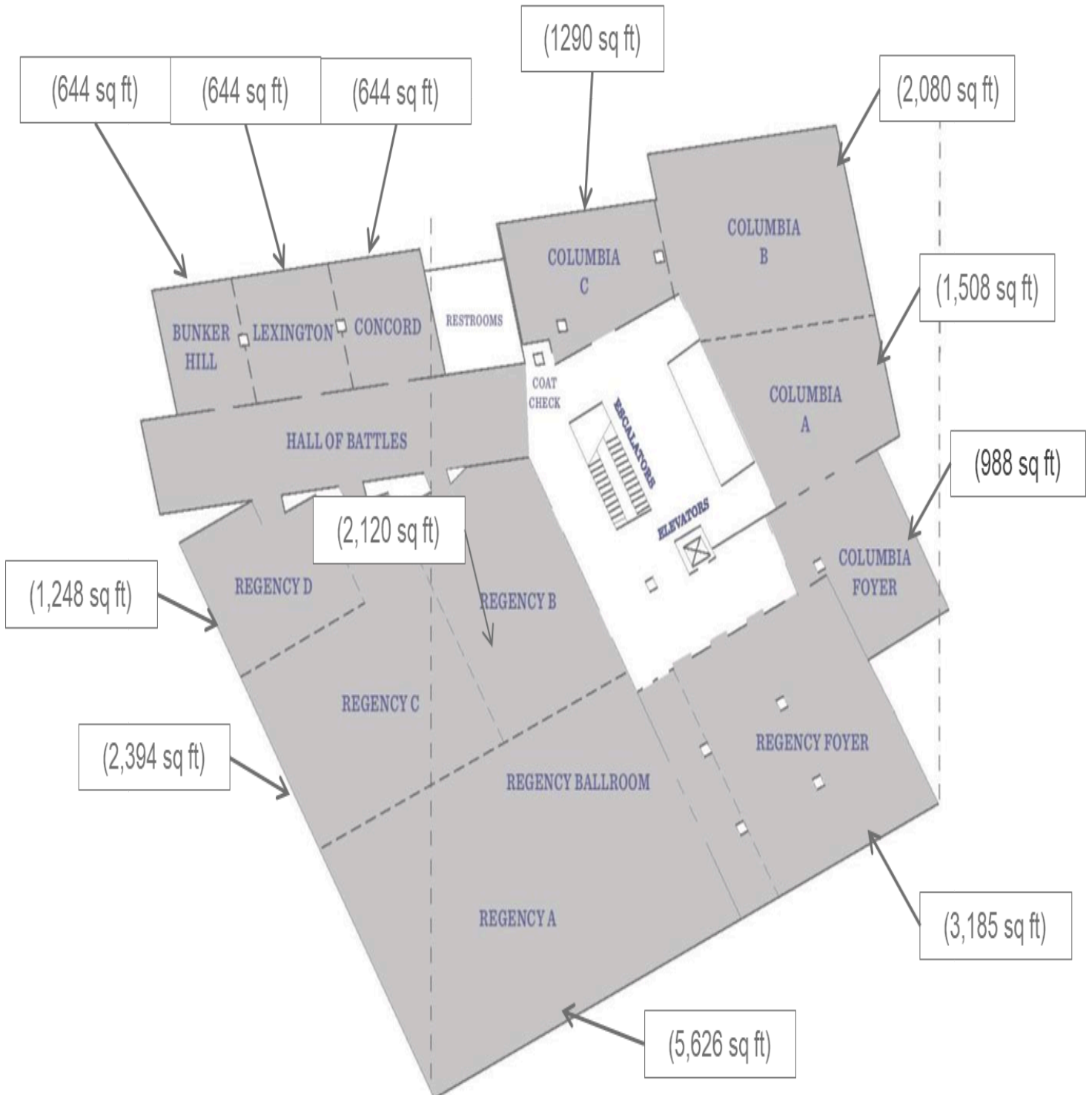|		
Workshop #22: 1st Workshop on Robust Machine Learning for Distribution Shifts https://sites.google.com/view/robustmlds24/home	Dec 15, Afternoon	Chair: Chen Zhao
Workshop #26: 8th International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2024) https://bda4cid.github.io/	Dec 15, Whole Day	Chair: Stephen McGough
Workshop: #31: International Workshop on Scalable and Deep Graph Learning and Mining https://project.inria.fr/sglm/	Dec 15, Whole Day	Wissem Inoubli
Dec. 16, Monday		
Workshop #16: 12th International Workshop on Distributed Storage and Blockchain Technologies for Big Data https://web.pkusz.edu.cn/itfna/workshop24/	Dec 16, Morning	Chair: Hong Tan
Workshop #41: Fourth International Workshop on Data science for equality, inclusion and well-being challenges (DS4EIW 2024) https://ds4eiw.github.io/4th-DS4EIW/	Dec 16, Morning	Chair: Tania Cerquitelli
Workshop #50: 2nd International Workshop on Creating Big Data for Morals, Ethics, and Values in Natural Language in an Automatically Inferable Format https://knowrel.com/IEEE_BIGDATA_2024_BIGETHICS/	Dec 16, Morning	Chair: Keisuke Nakamura
Workshop #27: 3rd International Workshop on Big Data Analytics for Health and Medicine (BDA4HM 2024) https://bda4hm.github.io/	Dec 16, Whole Day	Chair: Stephen McGough
Combined Workshops: #30 AI-Powered Renewable Energy Forecasting: Techniques and Challenges (AIPREF) & #32: AI-driven Agriculture: Opportunities and Challenges & #33: Cultural Heritage &BigData: Opportunities and Challenges https://sites.google.com/view/aipref2024/ https://sites.google.com/view/aida-2024/home-page https://sites.google.com/dimes.unical.it/chbdbigdata2024/home-page	Dec 16, Whole Day	Chairs: Ester Zumpano, Luciano Caroprese, Tommaso Ruga, Eugenio Vocaturo
Dec. 17, Tuesday		
Workshop #21: 5th International Workshop on User Understanding from Big Data https://u2bigdata.github.io/	Dec 17, Morning	Chair: Wutao Wei

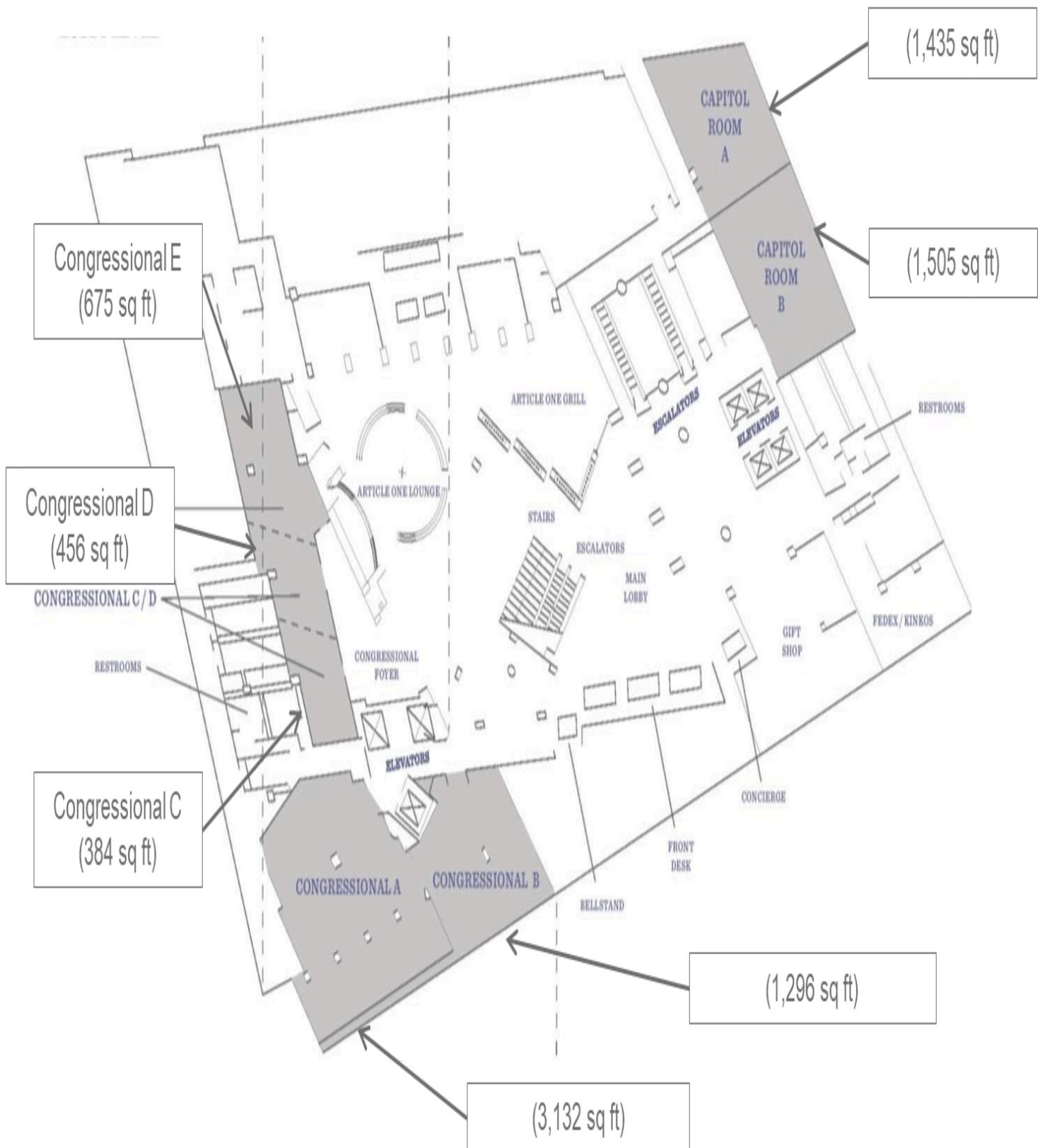|		
Workshop #29: Trustworthy Machine Learning for Healthcare https://sites.google.com/view/trustml4health24/	Dec 17, Morning	Chair: Xiao Shou
Workshop #28: 4th International Workshop on Big Data Analytics for Sustainability (BDA4S 2024) https://bda4s.github.io/	Dec 17, Whole Day	Chair: Stephen McGough

Online Challenge Cup Sessions

Dec. 15, Sunday		
Challenge Cup #5: Suicide ideation detection on social media Website: https://competitionpolyu.github.io/	Dec. 15, Morning	Chair: Jun Li 21118786r@connect.polyu.hk
Challenge Cup #4: Characterizing User Behavior in Social Networks: Propagation, Prediction, and Sensemaking Website: https://social-network-competition-gxqpe2ab8eh9atpe6hneu3.streamlit.app/	Dec. 15, Afternoon	Chair: Qingyun Sun sunqy@buaa.edu.cn
Dec. 16, Monday		
Challenge Cup #3: Building Extraction Generalization Challenge 2024 Website: https://www.kaggle.com/competitions/building-extraction-generalization-2024	Dec. 16, Morning	Chair: Shenglong Chen chen-sl@csis.u-tokyo.ac.jp
Dec. 17, Tuesday		
Challenge Cup #2: Optimized Road Damage Detection Challenge (ORDDC'2024) Website: https://orddc2024.sekilab.global/	Dec. 17 Whole Day	Chair: Deeksha Arya deeksha.arya1211@gmail.com
Challenge Cup #6: Challenges of Trustworthy AI in Distribution Shifts and Algorithmic Fairness Website: https://bigdatacup-e2j7tqwypbrkm8zsmqstmm.streamlit.app/	Dec. 17, Morning	Chair: Minglai Shao shaoml@tju.edu.cn

Hotel Floor Layout

Ballroom Level

(644 sq ft)

(644 sq ft)

(644 sq ft)

(1290 sq ft)

(2,080 sq ft)

(1,508 sq ft)

(988 sq ft)

(1,248 sq ft)

(2,120 sq ft)

(2,394 sq ft)

(3,185 sq ft)

(5,626 sq ft)

COLUMBIA C

COLUMBIA B

COLUMBIA A

COLUMBIA FOYER

BUNKER HILL

LEXINGTON

CONCORD

RESTROOMS

COAT CHECK

ESCALATORS

ELEVATORS

HALL OF BATTLES

REGENCY D

REGENCY B

REGENCY C

REGENCY FOYER

REGENCY BALLROOM

REGENCY A

Lobby Level

(1,435 sq ft)

(1,505 sq ft)

Congressional E
(675 sq ft)

Congressional D
(456 sq ft)

Congressional C
(384 sq ft)

CAPITOL ROOM A

CAPITOL ROOM B

RESTROOMS

ARTICLE ONE GRILL

ESCALATORS

ELEVATORS

ARTICLE ONE LOUNGE

STAIRS

ESCALATORS

MAIN LOBBY

GIFT SHOP

FEDEX / KINKOS

CONGRESSIONAL C / D

RESTROOMS

CONGRESSIONAL FOYER

ELEVATORS

CONCIERGE

FRONT DESK

CONGRESSIONAL A   CONGRESSIONAL B

BELLSTAND

(1,296 sq ft)

(3,132 sq ft)

2nd Floor Conference

Breakout
(475 sq ft)

GRAND
TETON

RESTROOMS

Everglades
(450 sq ft)

Breakout
(364 sq ft)

GLACIER

Yellowstone
(450 sq ft)

Breakout
(575 sq ft)

YOSEMITE

Breakout
(300 sq ft)

REDWOOD

Breakout
(575 sq ft)

BRYCE

SEQUOIA

GRAND
CANYON

Breakout
(340 sq ft)

OLYMPIC

Breakout
(494 sq ft)

Breakout
(546 sq ft)