Please note: The program is subject to change.

The schedule for keynote speakers may change slightly. | CST - Central Time Zone (Time in Lima City, Lima, Peru) | * Presenting author

Day 1 · Wednesday, October 28, 2026

TIME	TITLE	AUTHOR(S)
09h15 - 09h30	**Reception**	
09h30 - 10h00	**Welcome to SIMBig 2026** — Authorities of the UCSP and the Co-Chairs of SIMBig 2026	
Session 1: Applied NLP and Language Model Applications		
10h00 - 10h15	Where LLM Knowledge Belongs: A Controlled Study of Semantic Injection Surfaces in SASRec-Style Sequential Recommendation	Wizenfeld, Jeremy; Bouneffouf, Djallel*
10h15 - 10h30	Gender Bias Mitigation with Activation Steering and Fine-Tuning LLMs	Zhou, Hanqing*; Inkpen, Diana; Kantarci, Burak
10h30 - 11h30	*Invited Speaker:* **Dan JURAFSKY** ***Stanford University, USA***	Stanford University
	The Social Failures of Language Models as Conversational Partners People converse with language models for information, advice, and emotional support. In this talk I'll summarize studies in our lab showing that models fail in systematic ways as social interlocutors. We find that language models are socially sycophantic, linguistically overconfident, overly anthropomorphic, and epistemically self-centered. We then show that these flaws have real consequences for users: people interacting with models suffer consequences including overreliance, distorted judgment, and reduced personal responsibility. I'll discuss datasets and metrics, explore mitigations, and call for design, evaluation, and accountability mechanisms to protect user well-being.	
11h30 - 11h45	**Coffee Break** (15 min)	
11h45 - 12h00	Personality-Aware Transformer Features for Suicide-Note Analysis	Lazo Vasquez, Ricardo*; Ochoa Luna, José
12h00 - 12h15	Beyond Student Labels: Evaluating Prompt Sensitivity in AI-Generated Practice Problems for Multilingual Learners	Rudraraju, Satyanarayana*
12h15 - 12h30	Few-Shot Requirements Classification: An Empirical Evaluation of MPNet and BETO on English and Spanish Datasets	Mogollon, Sergio*
12h30 - 12h45	Semantic Review Mining for Regulatory Risk Detection	Chianumba, Ernest; Santoni, Jesse; Samuel, Raina*
12h45 - 13h00	Transformer-Based Models for Health Information Disorder Analysis: Misinformation Detection and Propagation Prediction	Alvarado Chavez, Margiory*; Manuico-Quequejana, Flavia; Pontalti Monari, Ana Carolina; Soriano-Vargas, Aurea
13h00 - 14h15	**Lunch** (1h 15 min)	
14h15 - 14h30	Transcription of Historical Documents using Seq2Seq Models	Laguna, Rodrigo*; Chavat, Felipe; Guillermo Moncecchi, Guillermo; Garat, Diego
14h30 - 14h45	Detecting and Mitigating Deception in Multilingual SLMs during Spanish-Language Tasks via Linear Probes	Sirena, Sergio*
Session 2: Image Processing and Computer Vision Applications		
14h45 - 15h00	Attention-Based Spatiotemporal Modeling for Crime Detection in Surveillance Videos	Túpac, Yván; Lazo Vasquez, Ricardo
15h00 - 16h00	*Invited Speaker:* **Gabriela CSURKA** ***NAVER LABS, France***	NAVER LABS Europe
	Towards 3D Foundational Models with Spatial intelligence In this talk I would like to take you on a journey through our development from CroCo to DUSt3R, which has brought a paradigm shift in 3D vision unifying the streamlined processing of 3D visual data. By exploring derivative and follow-up work from both our team and the broader community, I will highlight how this framework has opened new avenues for solving complex 3D challenges like map-free visual localization, robotic navigation, and beyond. Next, I will discuss it's extension towards different class of multi-view 3D reconstructions models using feed-forward transformer such as MUSt3R, Spann3R, CUT3R and VGGT enabling direct reconstruction of large, multi-view datasets. I'll continue with PanSt3R, a system built on MUST3R, that incorporates semantic understanding, allowing for joint 3D reconstruction and panoptic segmentation directly from unposed 2D image collections. By combining geometric reasoning with semantic perception, PanSt3R represents a significant step toward end-to-end 3D scene understanding from raw imagery. Finally, I will conclude my talk by discussing how to combine such 3D Foundational Models with LLMs to enhance their scene understanding and spatial intelligence.	
16h00 - 16h15	**Coffee Break** (15 min)	
16h15 - 16h30	Instance Segmentation of Defects in Hand-Spun Alpaca Fiber Yarn Using YOLO Models	Montoya Cubas, Carlos Fernando*; Hancco Champi, Fran Anthony; Mamani Meza, Johana Maria; Achahui Cruz, Milton Amed; Díaz Jurado, José Mauricio; Berrocal Cueto, Jhordan
16h30 - 16h45	From Monocular Video to 3D Pose and Kinematic Descriptors: A Padel-Oriented Pipeline with Limited 3D Data	Corral, Alejo*; Dominguez, Martin; Molina, Matias
16h45 - 17h00	Image–Physics Complementarity for M- and X-Class Solar Flare Prediction: A Multimodal Study with Swin3D and BiLSTM	Muñoz, Giomar*; Cardenas, Rolando

TIME	TITLE	AUTHOR(S)
17h00 - 17h15	Spatiotemporal Prediction of Land Use and Land Cover Change in Cusco Using Random Forest, Cellular Automata and Markov Chains	Choque-Navarro, Marcelo Fabian*; Medrano-Valencia, Ivan Cesar; Vera-Olivera, Harley; Alvarez-Mamani, Edwin
Session 3: Applied Machine Learning for Society		
17h15 - 17h30	Evaluating Climate Representations for Agricultural Productivity Modeling: Evidence from Peru	Urrea Rodriguez, Clara Patricia*; Villavicencio Lopez, Luis Felipe; Nuñez Del Prado Cortez, Miguel
17h30 - 17h45	Do TFMs Reduce the Need for Statistical Preprocessing in Small-Sample Excess-Zero Environmental Microbiome Data?	Orellana, Omar; Valenzuela, Luis; Marti, Luis; Chauvin, Gautier; Enaud, Raphael; Delhaes, Laurence; Avalos-Fernandez, Marta*
17h45 - 18h00	Latency-Aware Streaming Predictive Architecture for Silica Estimation in Iron Ore Flotation Using Machine Learning, Apache Kafka, and Digital-Twin-Style Monitoring	Puruguay, Sebastian*; Meza, Alexis; Vilcatoma, Joel; Calderon, Marks
18h00 - 18h15	Segment-Emergent Association Rules: Separating Temporal Signals from Partitioning Artifacts	Santisteban, Julio*; Arevalo Alvis, Jorge
18h15 - 18h30	Explainable Machine Learning and Spatial Equity Analysis of University Admission Outcomes in the Puno Region, Peru	Cutipa-Santi, Michael Newton*; Quelca-Velasquez, Daisy Crishyely; Cutipa-Santi, Yeni; Mendoza-Mollocondo, Charles Ignacio; Tito Lipa, José Pánfilo
18h30 - 18h45	3-Phase Multi-RAG Architecture: Optimization of Routing and Information Retrieval in University Admissions Processes	Arroyo-Paz, Antonio*; Chambi-Velasquez, Rommel; Hilacondo-Begazo, Andre

Day 2 · Thursday, October 29, 2026

TIME	TITLE	AUTHOR(S)
09h15 - 09h30	**Reception**	
Session 1: Health Applications and Biomedicine		
09h30 - 09h45	Machine Learning and Multi-Omic Enrichment Analysis for Breast Cancer Classification from mRNA and miRNA Microarray Profiles	Ramos Alvarez, Ismael Gerson*; Palma Ttito, Luis Beltran; Alzamora Paredes, Robert Wilbert; Rozas Huacho, Javier Arturo; Montoya Cubas, Carlos Fernando
09h45 - 10h00	Multimodal Brain Age Estimation with a beta-Variational Autoencoder and Gradient Boosting on a Multicentric Latin American Cohort	Barsotti, Damián*; Bazán, Nicolás M.; Della Bella, Gabriel
10h00 - 11h00	*Invited Speaker:* **Hoifung POON** ***Recursion, USA***	RECURSION
	Toward Virtual Patient: AI for Accelerating Biomedical Discovery Human biology stems from four billion years of evolution, which creates a proliferation of complexity and heterogeneity. The combinatorial explosion of biological hypotheses easily surpasses the number of atoms in the universe. Meanwhile, biomedical discovery advances at the snail pace of clinical trials, each taking years and costing over $100 million. As we enter the era of precision health and try to tailor treatments for each individual, continuing on today's discovery processes is clearly unsustainable. Our overarching research agenda lies in advancing multimodal generative AI to create a biomedical world model by bridging virtual patient and virtual cell, thus turning real-world patient journeys into a discovery engine through a generalized lab-in-the-loop flywheel. This enables us to synthesize population-scale real-world evidence from hundreds of millions of patients and accelerate biomedical discovery through AI-powered virtual clinical trials. This in turn fuels diverse applications from imminent productivity gains via automating digitized knowledge work, to moonshot creativity gains via creating biomedical wind tunnels for in silico simulation. By leveraging AI to close the loop across clinical care and frontier research, we have a clear line of sight toward scalable drug discovery.	
11h00 - 11h15	**Coffee Break** (15 min)	
11h15 - 11h30	Physics-Informed Graph Neural Networks for Missing Data Imputation in Unbalanced Three-Phase Distribution Grids	Peña Torres, Josemanuel Alejandro; Quispe Zavala, Rosmery Violeta; Gutierrez-Pachas, Daniel Alexis*
11h30 - 11h45	Improving Lightweight Models for Lung Cancer Detection via Two-Stage Knowledge Distillation	Navarro Guillen, Emerson*; Yauri Vidalon, Jose Elias; Vargas Sotaya, Yakelin; Zapata Casaverde, Richard; Fernández Jerí, Lissette Elvira; Luján Pérez, Diana
11h45 - 12h00	How Should Emergency Department Resource Utilization Be Clustered? A Comparative Study of Distance Metrics and Clustering Strategies	Elorga-Castagnet, Nadia*; Guerra-Adames, Ariel; Russon, Dylan; Doremus, Océane; Gil-Jardiné, Cédric; Avalos-Ferandez, Marta
12h00 - 13h00	*Invited Speaker:* **Hugo AERTS** ***Harvard University, USA***	HARVARD UNIVERSITY
	Beyond Prediction: Using AI to Discover Hidden Biology in Medical Images Artificial intelligence in medicine is often framed as a tool for automating clinical tasks or predicting patient outcomes. But its greater potential may be to reveal biological information that has remained hidden within routinely collected medical data. In this talk, I will discuss how advances in foundation models and multimodal AI are enabling medical images to be interrogated at scale, turning them into quantitative measurements of human biology. I will highlight recent work showing how AI analysis of routine CT imaging can measure thymic health in adults and uncover its relationships with cancer risk, survival, immune function, and response to cancer therapy. These studies illustrate a broader opportunity: moving beyond models that simply predict outcomes toward AI systems that generate biological insight, identify new biomarkers, and ultimately help guide prevention and treatment. I will discuss the opportunities and challenges of translating these discoveries into clinically useful tools and how this emerging paradigm could reshape the role of AI in biomedical research.	
13h00 - 14h00	**Lunch** (1 hour)	
14h00 - 14h15	Guidelines for the Design of an Artificial Intelligence-Based Mobile Application for Dietary Tracking and Gastrointestinal Pattern Recognition in University Students	Díaz Suárez, Jorge Eduardo*; Atencio, John; Frisancho, Sebastian; Iñiguez-Jarrin, Carlos
14h15 - 14h30	Machine Learning-Based Classification of Dengue Severity in Peru Using Public Health Data	Flores-Carbajal, Luz M.; Gonzales-Castillo, Julio C.; Mamani-Coaquira, Yonatan*
Session 2: Machine Learning Fundamentals, Semantic Search and Retrieval		
14h30 - 14h45	Retrieval-Augmented Fine-Tuning of Small Language Models for Domain-Specific Technical Reasoning in Peru's Sanitation Sector	Moreno Alvarado, Herman*; Ochoa-Luna, José
14h45 - 15h00	Adaptive Semantic Anchors: Persistent Geometric Regularization for Cold-Start Matrix Factorization	Ohayon, Ariel; Bouneffouf, Djallel*
15h00 - 16h00	*Invited Speaker:* **Omar ALONSO** ***Amazon Science, USA***	amazon
	Search in the LLMs Era LLMs are at the core of current advances in many information retrieval applications. Things are moving very fast, and it seems that techniques become outdated within weeks. Is the search problem solved? Are these new developments repetitions of older concepts, or a combination of both? Are we evaluating the right things? I'll present a personal perspective on these topics and outline the main challenges and opportunities.	
16h00 - 16h15	**Coffee Break** (15 min)	
16h15 - 16h30	Co-occurrence Graph and Relevance-Based Filtering for Precision Improvement in CSR-RAG Table Retrieval	Apaza Iruri, Mauricio*; Villanueva Llerena, Julissa
16h30 - 16h45	Contrastive Learning for Semantic Retrieval of Peruvian News Articles: A Comparative Evaluation Using a Journalistic Corpus	Arroyo-Paz, Antonio*; Ezcurra-Paima, Maria; Marroquin-Pinto, Paola
16h45 - 17h00	Offline Metrics for Explanations over Continuous Latent Factors	Espinoza, Edgar*; Sarmiento, Edgar

TIME	TITLE	AUTHOR(S)
17h00 - 17h15	A quantum benchmark case for Multi-swarm and quantum PSO	Ferrante, Agustin*; Dominguez, Martin; Osenda, Omar
Session 3: Machine Learning Fundamentals II		
17h15 - 17h30	The Carbon Footprint of Optimizers in Deep Neural Networks: A Green AI Benchmark	Camino, Adrian Emilio; Navarrete, Dax*; Flores-Reyes, Alejandro; Cabrera, Danny; Suntaxi, Gabriela; Recalde, Lorena; Martinez-Mosquera, Diana
17h30 - 17h45	A Model-Agnostic Temporal Explainability Framework for Financial Time Series Prediction	Zavalaga, Rushell*
17h45 - 18h00	Exploring Around the Collapse of Learned Heuristics with Restricted Novelty	Maldonado Quispe, Percy*; Gaona Briceño, Leonardo Gustavo
18h00 - 18h15	Privacy-Preserving Distributed Neural Networks	Cáceres Romero, Sergio; Candido, Renato; Silva, Magno*

Day 3 · Friday, October 30, 2026

TIME	TITLE	AUTHOR(S)
Session 1: Shared Task		
09h30 - 10h30	Shared Task on ASR and Language Resources for Indigenous Languages of South America	Huaman, Elwin*; Córdoba, Johana*
10h30 - 11h30	*Invited Speaker:* **Diana INKPEN** ***University of Ottawa, Canada***	uOttawa
	Legal Text Mining I will present an overview of legal text mining techniques based on LLMs, contextual embeddings, domain knowledge, and linguistic processing. I will discuss methods and results for several case studies such as: classifying court decisions, classifying court of appeal cases as allowed or not, classifying two cases as entailing each other or not, detecting the types of law violations, detecting named entities in legal texts, and classifying two trade agreement articles as being almost identical (stylistic differences only), similar but with relevant semantic differences, or unrelated.	
11h30 - 11h45	**Coffee Break** (15 min)	
Session 2: Data-Driven Systems and Intelligent Applications		
11h45 - 12h00	Towards Reactive and Intelligent Smart Laboratories: A Multi-Agent Context-Aware Software Architecture	Paucar Curasma, Herminio*; Cezar Estrella, Julio
12h00 - 12h15	Scaling Character Adherence: Efficiently Adapting Small Language Models for Interactive NPCs	Butron Cuayla, Daryl*; Lazo Vasquez, Ricardo Manuel; Ochoa Luna, José Eduardo
12h15 - 12h30	Semantic Search of Scientific Abstracts through HNSW Indexing in Vector Databases	Arroyo-Paz, Antonio*; Callocondo-Garay, Santiago; Mejia-Ramos, Piero
12h30 - 12h45	When Semantic Alignment Helps Sequential Recommendation	Wexler, David; Bouneffouf, Djallel*
12h45 - 13h00	Face Verification and Presentation Attack Detection for University Access Control	Vargas Saldaña, Daniella*
13h00 - 14h00	**Lunch** (1 hour)	
14h00 - 14h15	Data-Driven Geostatistical Optimization of LoRaWAN Sensor Placement for Urban Heat Island Monitoring in Piura, Peru	Ocaña Rufino, Jazmín; Samillán Vera, Ángel; La Rosa Lama, Gerson*; Soto Bohórquez, Juan
Session 3: Education, Society and Machine Learning Applications		
14h15 - 14h30	Operationalizing Bias Mitigation in Scholarly Information Systems: Fairness-Utility Trade-offs across Prediction and Retrieval	Recalde, Lorena*; Sarango, César; Martinez-Mosquera, Diana; Suntaxi, Gabriela
14h30 - 14h45	Sentiment Analysis in Collao Quechua under Limited Data: Lessons from an Automatically Annotated Corpus	Marín-Rodríguez, Sintia Vallet; Paccotacya-Yanque, Rosa*
14h45 - 15h00	Digital Reading and Writing Indicators for Early Dyslexia Risk Screening in Primary School Students: A Conceptual Mobile Model	Campos, Silvia*; Huamani, Joseph; Quispesivana, Claudio
15h00 - 15h15	What Survives an Uneven Record? Fine-Grained Crime Prediction under Reporting Bias: Evidence from Peru	Vilchez, Rody*; Velasquez, Christian; Alarcon, Carlos
15h15 - 15h30	**Coffee Break** (15 min)	
15h30 - 15h45	Explainable Credit Risk Assessment with FT-Transformer and Multi-Agent LLMs	Maldonado Quispe, Percy*; Huanca Olazabal, Cristhian David
15h45 - 16h00	Unsupervised Learning for Profiling University Students with Parental Responsibilities	Cutipa-Santi, Yeni; Cutipa-Santi, Michael Newton*; Quelca-Velasquez, Daisy Crishyely; Apaza-Cutipa, Renzo; Ibañez-Quispe, Vladimiro
16h00 - 16h15	**SIMBig 2026 Closing** — Authorities of the UCSP and the Co-Chairs of SIMBig 2026	