# GENERAL PROGRAM - SIMBig 2024

## Novembre 20, 2024

| Hour | Title | Speaker |
|---|---|---|
| 09h30 - 11h00 | Reception | |
| 11h00 - 11h15 | Welcoming to SIMBig2024 | Authorities of the Universidad Nacional de Moquegua and the Co-Chairs of SIMBig 2024 |
| **Session 1** | | |
| 11h15 - 11h30 | Explainable Sentiment Analysis on Restaurant Reviews using an Evolutionary Algorithm | Edward Hinojosa-Cardenas, Leticia Laura-Ochoa, Edgar Sarmiento |
| 11h30 - 12h30 | *Invited Speaker:*<br>**Anna Korhonen**<br>**University of Cambridge, UK**<br><br>Title: Towards globally equitable language technologies<br><br>Abstract: Language technologies can now offer effective support to communication, education, healthcare, and many other aspects of human life. Yet, these technologies are not distributed equally. They are only available for a small part of the world's 7.9 billion population, mainly those living in the Global North. This is because the resources needed for high quality language processing are limited or lacking for the vast majority of the world's over 7,000 living languages. In this talk I will discuss the scientific and socioeconomic implications of this situation, along with the many methodological challenges that we must overcome. My talk will outline an approach that could transform the way we approach multilingual NLP, and substantially improve our understanding of how language technologies can be made fair and inclusive at a global level. | |
| 12h30 - 12h45 | Coffee Break | |
| **Session 2** | | |
| 12h45 - 13h00 | Making an Under-Resourced Language available on the Wikidata Knowledge Graph: Quechua Language | Elwin Huaman, Jorge Luis Huaman, Wendi Huaman |
| 13h00 - 13h15 | Predictive model applying sentiment analysis on Tweets to determine the behavior of the cryptocurrency Bitcoin | Henry Condori Alejo, Ernesto Zhildeer Chura-Flores, Guina Alzamora Sotomayor |
| 13h15 - 13h30 | ThreatGram 101 - Extreme Telegram Replies Data with Threat Levels | Kamalakkannan Ravi, Jiann-Shiun Yuan |
| 13h30 - 14h30 | Lunch | |
| **Session 3** | | |
| 14h30 - 14h45 | Comparison of Recurrent Neural Network Architectures to Answer Questions of the Universities Admission Process | Maribel E. Coaguila, Mariela M. Nina, Yoselin Arocutipa, Estrella Velásquez, Jesus Rocca, Honorio Apaza |
| 14h45 - 15h00 | Enhancing Academic Profiling with Advanced NLP Techniques: SBERT + KeyBERT | Carlos A. Ayala-Tipan, Lorena Recalde, Gabriela Suntaxi |
| 15h00 - 16h00 | *Invited Speaker:*<br>**Aline Villavicencio**<br>**University of Exeter, UK**<br><br>Title: Investigating Idiomaticity in Language Models: or going after untamed language models with a cause<br><br>Abstract Large language models have been successfully used for capturing distinct (and very specific) word usages, and therefore could provide an attractive alternative for accurately determining meaning in language. However, these models still face a serious challenge when dealing with non-literal language, like that involved in Multiword Expressions (MWEs) such as idioms (make ends meet), light verb constructions (give a sigh), verb particle constructions (shake up) and noun compounds (loan shark). MWEs are an integral part of the mental lexicon of native speakers often used to express complex ideas in a simple and conventionalised way accepted by a given linguistic community. Determining whether they represent a challenge for current models, and accurately modelling them have huge potential for improving the precision, naturalness and fluency in interpretation and generation tasks like machine translation. In this talk, I will present an overview of how advances in models have made an impact in performance for idiomaticity and MWE handling. I will concentrate on what models seem to incorporate of idiomaticity, as idiomatic interpretation may require knowledge that goes beyond what can be gathered from the individual words of an expression (e.g. “dark horse” as an unknown candidate who unexpectedly succeeds). |  |
| 16h00 - 16h15 | Coffee Break | |
| 16h15 - 16h30 | Leveraging LLMs for Enhanced Personality Trait Classification | Giulia Naval, Oscar Linares, José Eduardo Ochoa Luna |
| 16h30 - 16h45 | EPIC: Enhancing Privacy through Iterative Collaboration | Prakash Chourasia, Heramb Lonkar, Sarwan Ali, Murray D. Patterson |
| 16h45 - 17h00 | Detection of Polarizing Narratives in Social Media through Machine Learning during Political Unrest | Angel Javier Quispe-Carita, Renzo Apaza-Cutipa, Juan Carlos Juarez-Vargas, Milton Antonio Lopez-Cueva, Ernesto Nayer Tumi-Figueroa, Fred Torres-Cruz |
| 17h00 - 17h15 | DWFL: Enhancing Federated Learning through Dynamic Weighted Averaging | Prakash Chourasia, Tamkanat . Ali, Sarwan Ali, Murray D. Patterson |
| 17h15 - 17h30 | Enhancing Bert Classification by Increasing Training Corpus Using Paraphrasing | Carlos Atencio Torres, Alvaro Mamani-Aliaga, Julio A. Vera, Pablo Calcina-Ccori |

## Novembre 21, 2024

| Hour | Title | Speaker |
|---|---|---|
| 08h30 - 09h00 | Reception | |
| **Session 4** | | |
| 09h00 - 09h15 | An Scoping Review: Explainable Deep Learning Approach for Automated Sismo-Volcanic Event Classification | Edison G Paria Fernandez, Edgar Sarmiento, Reynaldo Alfonte Zapana |
| 09h15 - 10h15 | *Invited Speaker:*<br>**Andrew Gentles**<br>**Stanford University, USA** |  |
| | "Title: Large scale analysis of cancer cohorts and single cell atlases to identify cell states and ecosystems associated with immunotherapy responses<br><br>Abstract: Tumor cellular composition and spatial organization play critical roles in influencing treatment responses and survival outcomes in cancer. New biomarkers and therapeutic strategies are needed, particularly in the context of immunotherapy where resistance can lead to patient relapse even after impressive initial responses. Recently we introduced EcoTyper, a computational framework that integrates single cell RNA sequencing with bulk RNA-seq enabling the identification of cancer-specific cell types and their association with patient outcomes. Pancancer analysis with EcoTyper revealed specific cell states and associations that were strongly associated with immunotherapy response. We now extend our approach to cancer-specific identification of the cellular and molecular determinants of immunotherapy response by integration of patient cohorts annotated with therapy responses, with single cell atlases for individual cancer types. We specifically identify cell states that preferentially express response-associated genes, and map these to spatial localization. This work aims to uncover new therapeutic targets and biomarkers, ultimately enhancing treatment strategies through a deeper understanding of tumor ecosystems and the interplay of cellular interactions in cancer progression." | |
| 10h15 - 10h30 | A Deep Learning Model for Ingredient and Meal Quantity Estimation in Type 2 Diabetes Care | Sergio Andres Flores Ñahuis, Renzo Arturo Paredes Villagra and Luis Martín Canaval Sánchez |
| 10h30 - 10h45 | Preserving Hidden Hierarchical Structure: Poincaré Distance for Enhanced Genomic Sequence Analysis | Sarwan Ali, Haris Mansoor, Prakash Chourasia, Imdad Ullah Khan, Murray D Patterson |
| 10h45 - 11h00 | Hybrid Quantum Model for Brain Tumor Classification | Sergio L. Ramos, Carlos Atencio Torres, José Eduardo Ochoa Luna |
| 11h00 - 11h15 | Coffee Break | |
| **Session 5** | | |
| 11h15 - 11h30 | Prediction of maximum rainfall for Early warning of heavy rainfall with Multivariable Long-term memory networks and Temporal Transformers | Yhon Fuentes Huaman, Kristhel Calderon Aedo, Vanesa Salas Peña, Anjhela Rosa Callo Mamani, Alberto Callo Yucra, Claudio Huancahuire Bravo, Elena Gonzalez Mamani, Jessica Alvarado Ramos, Guido Bravo Mendoza |
| 11h30 - 12h30 | *Invited Speaker:*<br>**Aaron Courville**<br>**Université de Montréal, Canada** | Université de Montréal |
| | Title: Multi-agent interactions: Why can't they just get along?<br><br>Abstract: In many real-world scenarios, agents optimizing their own utility will interact with a dynamic mix of cooperation and competition. Despite the ubiquitous relevance of such settings, decentralized machine learning algorithms have struggled to find equilibria that maximize individual utility while preserving social welfare. In this talk I will discuss our latest efforts in this direction by introducing two algorithms: (1) Q-value Shaping and (2) Advantage Alignment, both are novel decentralized RL algorithm tailored to optimizing an agent's individual utility while fostering cooperation among adversaries in partially competitive environments. The central innovation with both of these algorithms is the assumption that, during training, the opponent samples actions proportionally to their action-value function Q. Experimental results demonstrate the effectiveness of our approach at achieving state-of-the-art performance in benchmark scenarios such as the Iterated Prisoner's Dilemma and the Coin Game. We believe this method to be an important step toward training agents for practical multi-agent applications. | |
| 12h30 - 12h45 | Predicting Shear Strength at the Joint of a Beam and a Column Using Supervised Learning and Multivariate Neural Networks | Yhon Fuentes Huaman, Yessenia Yari Ramos, Raymundo Aguirre Ramos |
| 12h45 - 13h00 | An efficient automatic classification hibryd model to identify images of commercial starchy corn | Diego Alonso Dorado Torres, Yeshica Isela Ormeño Ayala, Lauro Enciso Rodas |
| 13h00 - 14h30 | Lunch | |
| **Session 6** | | |
| 14h30 - 14h45 | Transformer GPT-2 Model for Code Prediction in Machine Learning Projects | Tito Peng Wu Pan and Luis Martín Canaval Sánchez |
| 14h45 - 15h45 | *Invited Speaker:*<br>**Krishnaram Kenthapadi**<br>**Oracle Health, USA** | ORACLE |
| | Title: Deploying Trustworthy Generative AI<br><br>Abstract: While generative AI models and applications have huge potential across different industries, their successful commercial deployment requires addressing several ethical, trustworthiness, and safety considerations. These concerns include domain-specific evaluation, hallucinations, truthfulness and grounding, safety and alignment, bias and fairness, robustness and security, privacy, unlearning, and copyright implications, calibration and confidence, and transparency. In this talk, we first motivate the need for adopting responsible AI principles when developing and deploying large language models (LLMs), text-to-image models, and other generative AI models, and provide a roadmap for thinking about responsible AI and AI observability for generative AI in practice. Focusing on real-world generative AI use cases (e.g., evaluating LLMs for robustness, security, bias, etc. especially in health AI applications and user-facing & enterprise-internal chatbot settings), we present practical solution approaches / guidelines for applying responsible AI techniques effectively and discuss lessons learned from deploying responsible AI approaches for generative AI applications in practice. This talk will be based on our KDD'24 LLM grounding and evaluation tutorial and ICML/KDD/FAccT 2023 trustworthy generative AI tutorial. | |
| 15h45 - 16h00 | Spanish Historical Handwritten Text Recognition with Deep Learning | Gustavo Jorge Choque Dextre, Ferdinand Pineda |
| 16h00 - 16h15 | Development of a convolutional neural network model to improve quality control of Hass avocado in the Paucartambo Agricultural Valley, Pasco, Perú | Arnold Anthony Huaman, Luis Andre Cosco Escobedo, David Francisco Leva Mescco |
| 16h15 - 16h30 | Coffee Break | |
| **Session 7** | | |
| | *Invited Speaker:*<br>**Mona Diab**<br>**Caneguie Mellon University, USA** | Carnegie Mellon University |

| | | |
|---|---|---|
| 16h30 - 17h30 | Title: Beyond Simple Accuracy: Building Culturally Aligned, Transparent, and Human-Like AI for a Responsible AI/NLP Future<br><br>Abstract: As large language models (LLMs) become integral to various domains, the need for responsible, culturally sensitive, and human-aligned AI has never been more pressing. This talk explores a multi-faceted approach to advancing ethical and effective LLMs, drawing on recent research in cultural alignment, transparency, bias mitigation, professional standards, and personality modeling. By combining such insights, we can develop LLMs that not only meet technical standards but also align with the values, expectations, and diverse needs of their users. This keynote will inspire AI practitioners to consider these dimensions as essential pillars for the next generation of language models. | |
| 17h30 - 17h45 | Methodological Approach for the Development of a Mountain Glacier Dataset Using Satellite Imagery for Deep Learning Applications | William Condori Quispe, Jose Herrera Quispe, Oscar Jimenez Flores, Juan Jimenez Castilla |
| 17h45 - 18h00 | Parallel Multi-objective Evolutionary Algorithms: A Systematic Literature Review | Saulo Gustavo Machicao-Mollocondo, Edward Hinojosa-Cardenas |
| 18h00 - 18h15 | Safycash: Mobile Application for Authentic Banknote Detection Using Image Processing Libraries and Convolutional Neural Networks | Erick A. Quispe Bernardo |

# Novembre 22, 2024

| Hour | Title | Speaker |
|---|---|---|
| 08h30 - 09h00 | Reception | |
| **Session 8** | | |
| 09h00 - 09h15 | Small Electra for Hate: A Transformer Lightweight Approach to Hate Text Classification | Ricardo Manuel Lazo Vasquez, José Eduardo Ochoa Luna |
| 09h15 - 10h15 | *Invited Speaker:*<br>**Huan Liu**<br>**Arizona State University, USA** |  |
| | Title: Embracing Generative AI for Social Media Mining<br><br>Abstract: In this talk, we will present emerging opportunities for researchers in AI and Big Data, through the lens of social media. Generative AI, in particular, Large Language Models (LLMs), has added new challenges. We use examples to illustrate (1) fundamental problems associated with multi-modal data like social media, challenging common practice and existential understanding in machine learning and data mining, (2) how we embrace the power of LLMs to solve perplexing problems, and (3) developing novel algorithms for responsible LLMs. Seeking interdisciplinary collaborations, we contemplate the promising future of data science and data mining in the rapid development of AI. | |
| 10h15 - 10h30 | Strategies to Mitigate University Dropout Using Classification Algorithms in Engineering Programs | Tania Y. Anahua, Yana Mamani |
| 10h30 - 10h45 | Electricity Demand Time Series Benchmark | Miguel Nunez-del-Prado, Juan Carlos Orihuela-Solis |
| 10h45 - 11h00 | Application of Multilayer Perceptron: Analysis of socioeconomic factors of poverty in Peru | Almir Vargas Mamani |
| 11h00 - 11h15 | Coffee break | |
| **Session 9** | | |
| 11h15 - 11h30 | Personality Traits and Voter Intentions: A Computer Vision and Deep Learning Study of Simulated Peruvian Elections in 2022 | Alejandro Aybar-Flores |
| 11h30 - 11h45 | Enhancing FAIRness of Spatiotemporal Data through Semantics: A Case Study with EarthPortal and EaSy Data | Christelle Pierkot, Hélène Bressan, Guillaume Alviset |
| 11h45 - 12h00 | A Comparison of Classification Algorithms for Disaster Tweets | Usman Khan |
| 12h00 - 12h15 | Development of 1D-CNN Methods for Classifying Pediatric Epilepsy through EEG Signals | Oscar Flores-Palermo, Christian Espiritu-Cueva, Willy Ugarte |
| 12h15 - 13h00 | Clossing of SIMBig 2022 | Authorities of the Universidad Nacional de Moquegua and the Co-Chairs of SIMBig 2024 |