

# Curriculum Vitae    Esaú Villatoro-Tello, Ph.D.

## Research Focus

- State-of-the-art Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG): Impact and applications in conversational data.
- Natural Language Processing and Mental health: AI applied in Clinical Psychology, Cognitive and Behavioral Processes analysis.
- Spoken and Natural Language Understanding (SLU/NLU): Intent detection and slot filling in agent-client conversations
- Natural Language Processing (NLP) and Text and Data Mining/Analysis: Authorship Analysis and Profiling
- Computational Linguistics (CL): Automation of Linguistic Theory
- Ethical Artificial Intelligence: Intelligible Models

## Qualifications and Experience

*Position:*                    **Researcher**  
*Institution:*                Idiap Research Institute  
*Address:*                    Rue Marconi 19, CH-1920 Martigny, Switzerland.  
*e-mail:*                      [esau.villatoro@idiap.ch](mailto:esau.villatoro@idiap.ch); *URL:* <https://www.idiap.ch/~evillatoro/>  
*Period:*                      2021 (November) – present

*Position:*                    **Academic Visitor**  
*Institution:*                Idiap Research Institute  
*Address:*                    Rue Marconi 19, CH-1920 Martigny, Switzerland.  
*Period:*                      2019 (September) – 2021(October)

*Position:*                    **Full-Time Researcher/Professor (Level C)**  
*Institution:*                Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C)  
*Address:*                    Avenida Vasco de Quiroga 4871, Colonia Santa Fé Cuajimalpa. Delegación Cuajimalpa de Morelos, C.P. 05348, México City, Mexico.  
*Period:*                      2012 (June) – 2022 (November)

*Position:*                    **Research Scientist (Postdoc)**  
*Institution:*                Mexican Petroleum Institute (IMP), Mexico  
*Address:*                    Eje Central Lázaro Cárdenas Norte 152 San Bartolo Atepehuacan C.P. 07730, Mexico City, Mexico  
*Period:*                      2010 (September) – 2012 (June)

*Degree:*                      **Research Assistant (Ph.D.) in Computer Science**  
*Institution:*                National Institute of Astrophysics Optics and Electronics (INAOE)  
*Thesis:*                      *Sample-based re-ranking in Geographic Information Retrieval*

*Address:* Luis Enrique Erro # 1, Tonantzintla, Puebla, Mexico C.P. 72840  
*Period:* 2007 (March) – 2010 (October)

*Degree:* **Master's in Computer Science (MSC)**  
*Institution:* National Institute of Astrophysics Optics and Electronics (INAOE)  
*Thesis:* *Single and Multiple Text Summarization*  
*Address:* Luis Enrique Erro # 1, Tonantzintla, Puebla, Mexico C.P. 72840  
*Period:* 2004 (August) – 2007 (February)

*Degree:* **Bachelor's in computer science (BSC)**  
*Institution:* Benemerita Universidad Autonoma de Puebla (BUAP), Mexico  
*Address:* 14 Sur esq. San Claudio. Col. San Manuel, Puebla, Puebla, México  
*Period:* 1999 (July) – 2004 (May)

## Funded Research Projects

### *As Principal Investigator (PI)/Project Leader*

1. **ORIENTER: tOwards undeRstanding and modellIng the language of mENTAL health disordERs.**  
*To be developed in:* Idiap-CH, CIMAT-MX, and INPRFM-MX  
*Period:* 2024-2028 (Kick-off meeting TBD)  
*Project reference/Sponsor:* SNSF-SPIRIT Program 2023, project id IZSTZ0\_223488 / 1, CH  
*Type of project:* Fundamental research.  
*Goal:* To investigate what aspects of patients' language and behaviour can be effectively and efficiently modelled by very recent AI techniques in the diagnostic construct of depression disorders.  
*My role:* (PI) main researcher and project coordinator of the consortium activities.
2. **Automatic Personality Traits Recognition using multimodal features** (*Representación Multimodal para la Categorización Automática de Rasgos de Personalidad*)  
*Developed in:* Information Technologies Department, Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.  
*Period:* May 2017- September 2019  
*Project reference/Sponsor:* CONACyT Research Program 2015. Project number: CB-2015-01-258588, Mexico.  
*Type of project:* Fundamental research.  
*Goal:* Evaluate the impact of multimodal features in the task of Personality traits recognition. Particularly, textual features in combination with attributes extracted from the images of handwritten texts are being considered. In parallel, a large multi-modal Spanish corpus was constructed.  
*My role:* (PI) Main researcher, project coordinator and supervisor of the involved students.
3. **Platform of Bioinformatic Analysis for Genomic Research and Diagnosis. Phase 3: incorporation of Artificial Intelligence algorithms and analysis of structural variants** (*Plataforma Integral de Análisis Bioinformático para el Diagnóstico e Investigación Genómica. Fase 3: incorporación de algoritmos de Inteligencia Artificial y análisis de variantes estructurales*)  
*Developed in:* Information Technologies Department, Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico in collaboration with Web and Internet and Network Technology for Enterprise Resources (WINTER) SA de CV.  
*Period:* February 2018- January 2019  
*Project reference/Sponsor:* Support for Innovation Program (PEI-CONACyT) 2018, project number 253441. Mexico.

*Type of project:* Industrial project

*Goal:* Development of an Information Extraction system specialized in finding clinical information associated with genetic variants in scientific papers. Natural Language Processing methods in combination with machine learning algorithms were employed.

*My role:* (PI) Main researcher, project coordinator and supervisor of the involved students and researchers.

4. **International Workshop on Computational Linguistics: Towards the deep analysis of digital documents** (*Taller Internacional de Lingüística Computacional: hacia el análisis profundo de documentos digitales*)

*Developed in:* Information Technologies Department, Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico in collaboration with IIMAS-UNAM, CIC-IPN, UNAL-Colombia.

*Period:* June 2019

*Project reference/Sponsor:* Programa de Apoyos para Actividades Científicas, Tecnológicas y de Innovación, Dirección Adjunta de Desarrollo Científico 2018 CONACyT. Project number 293097, Mexico.

*Type of project:* Science Divuligation and Communication.

*Goal:* To provide a training event with a global scope aiming at updating participants about the most recent advances in the critical and fast-developing area of deep learning and its impact on Natural Language Processing. All the material from this workshop can be found at <http://dti.cua.uam.mx/tallerPLN2018/>

*My role:* (PI) Main researcher and coordinator of the event.

5. **Cloud bioinformatics platform for the identification of genetic variants, an application for molecular diagnosis and genomic research** (*Plataforma bioinformática en nube para la identificación de variantes genéticas, una aplicación para el diagnóstico molecular y la investigación genómica*)

*Developed in:* Information Technologies Department Universidad Autónoma Metropolitana Unidad Cuajimalpa in collaboration with Web and Internet and Network Technology for Enterprise Resources (WINTER) SA de CV., Mexico.

*Period:* February 2017- January 2018

*Project reference/Sponsor:* Support for Innovation Program (PEI-CONACyT) 2017, project number 242211, Mexico.

*Type of project:* Industrial project.

*Goal:* Develop a machine-learning approach for the identification of genetic variants with pathogenic potential.

*My role:* (PI) Main researcher, project coordinator and supervisor of the involved students and researchers.

6. **Sexual Predators Identification in Chat Conversations Employing Sequential Information** (*Detección de Depredadores Sexuales en Conversaciones de Chat con Información Secuencial*)

*Developed in:* Information Technologies Department, Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.

*Period:* September 2013- December 2014

*Project reference/Sponsor:* New Full-time Professors Incorporation Program SEP-PROMEP 2013. *Project number:* SEP-PROMEP-48510349, Mexico.

*Type of project:* Fundamental research

*Goal:* Develop a novel method for automatically identifying sexual predators in chat conversations based on text representations that incorporate sequential data.

*My role:* (PI) Main researcher, project coordinator and supervisor of the involved students.

*As collaborator/co-applicant*

7. **Multilingual and Cross-cultural interactions for context-aware, and bias-controlled dialogue systems for safety-critical applications (ELOQUENCE Project)**

*Developed in:* Idiap Research Institute, Project Coordinator: Telefónica I+D, Spain.

*Period:* 2024 – 2026 (Kick-off meeting January 2024)

*Project reference/Sponsor:* Project recently funded under the HORIZON-CL4-2023-HUMAN-01-03: Natural Language Understanding and Interaction in Advanced Language Technologies (AI Data and Robotics Partnership) (RIA) call. EU.

*Type of project:* Research and Innovation.

*Goal:* ELOQUENCE's key objective is to better comprehend unstructured dialogues and translate them into explainable, safe, knowledge-grounded, trustworthy and bias-controlled language models. ELOQUENCE is an elite alliance of 16 partners across 10 countries. We bring together top R&D centres, academic leaders, and conversational AI innovators, focusing on groundbreaking advancements in natural language processing, robotics, and cognitive systems.

*My role:* Collaborator - I actively worked on defining the methodology and specifying the goals, activities, key performance indicators, and deliverables for various work packages (WPs). I was in charge of coordinating the activities involved in the definition of WP3 (Explainability, automated reasoning & context-based interaction in unstructured dialogues) of the ELOQUENCE proposal.

8. **Design of materials for large-scale electrochemical energy storage, based on “synthesizability” and chemical space of abundant species in Mexican mining.**

*Developed in:* Universidad Autonoma Metropolitana Unidad Iztapalapa

*Period:* 2023 - 2026

*Project reference/Sponsor:* Frontier Research Program 2023 (Ciencia de Frontera 2023) Project number: CF-2023-G-1266, Mexico

*Type of project:* Fundamental Research.

*Goal:* The primary objective of the project is to generate a search protocol for new materials for intercalation battery electrodes.

*My role: (Co-PI) Co-applicant:* I helped to define the NLP technologies that would be applicable to the posed problem, in particular, information extraction in very narrow and low-resource domains.

9. **Uniphore-Idiap Partnership. Industrial collaboration on data analytics for call centres. (Industrial project)**

*Developed in:* Idiap Research Institute

*Period:* 2020 – present

*Project reference/Sponsor:* Uniphore Software Systems Inc., Palo Alto, CA, USA

*Project type:* Industrial project.

*Goal:* Development of speech and natural language processing systems for Uniphore (industrial partner based in USA and India)

*My role:* Collaborator & Team leader (Group of 5 persons). I am responsible for coordinating the activities of Idiap's team, consisting of five members, and serve as the lead researcher on SLU/NLU and LLMs technologies.

10. **Real-time network, text, and speaker analytics for combating organized crime (ROXANNE Project)**

*Developed in:* Idiap Research Institute

*Period:* 2020 –2022

*Project reference/Sponsor:* Horizon 2020 - European Commission. EU

*Project type:* Research and Innovation

*Goal:* Develop a platform to aid police investigations by integrating multiple technologies: speaker recognition, ASR, NER, Topic detection, Coreference resolution, and network analysis

*My role:* Collaborator - Advisor. I actively participated in the NLP team meetings, focusing on Named Entity Recognition, Topic Detection and Tracking, and Author Profiling activities.

11. **SARAL: Summarization and domain-adaptive Retrieval of Information Across Languages**

*Developed in:* Idiap Research Institute

*Period:* 2020 – 2022

*Project reference/Sponsor:* The Consortium was a part of the [IARPA MATERIAL project](#). USA.

*Project type:* Research.

*Goal:* To develop systems to support English queries on low-resource, non-English languages (e.g. Tagalog and Swahili).

*My role:* Collaborator - Main researcher working on the NLP and Information Retrieval tasks.

## Publications

### *Refereed Journal Publications (Indexed by Journal Citation Reports)*

1. Natural Language Understanding for Navigation of Service Robots in Low-Resource Domains and Languages: Scenarios in Spanish and Nahuatl. Amadeo Hernández, Rosa M. Ortega-Mendoza, **Esau Villatoro-Tello**, César J. Camacho-Bello, Obed Pérez-Cortés. In *Mathematics*. 2024; 12(8):1136. April 2024
2. A lexical-availability-based framework from short communications for automatic personality identification. Gabriela Ramírez-de-la-Rosa, Héctor Jiménez-Salazar, **Esau Villatoro-Tello**, Verónica Reyes-Meza & Jaime Rojas-Avila. In *Cognitive Systems Research Journal*, Vol. 79. January 2023
3. Applying Attention-Based Models for Detecting Cognitive Processes and Mental Health Conditions. **Esau Villatoro-Tello**, Shantipriya Parida, Sajit Kumar, and Petr Motliceck. In *Cognitive Computation Journal*, Springer, 2021. July 2021
4. Predicting consumers engagement on Facebook based on what and how companies write. Érika S. Rosas-Quezada, Gabriela Ramírez-de-la-Rosa, and **Esau Villatoro-Tello**. In *Journal of Intelligent & Fuzzy Systems (JIFS)*, vol. 39, no. 2, pp. 2365-2377. June 2020.
5. A Comparative Analysis of Distributional Term Representations for Author Profiling in social media. Álvarez-Carmona, M. Á., **Villatoro-Tello, E.**, Montes-y-Gómez M., and Villaseñor-Pineda, L. In *Journal of Intelligent & Fuzzy Systems (JIFS)*, vol. 36, no. 5, pp. 4857-4868, May 2019.
6. A visual approach for age and gender identification on Twitter. Luis Pellegrin, Miguel A. Alvarez-Carmona, Manuel Montes-Y-Gomez, Fernando Sánchez-Vega, Hugo Jair Escalante, A. Pastor López-Monroy, Luis Villaseñor-Pineda and **Esau Villatoro-Tello**. In *Journal of Intelligent & Fuzzy Systems (JIFS)*, vol. 34, no. 5, pp. 3133-3145, May 2018.
7. TxPI-u: A Resource for Personality Identification of Undergraduates. Gabriela Ramírez-De-La-Rosa, **Esau Villatoro-Tello** and Héctor Jiménez Salazar. In *Journal of Intelligent & Fuzzy Systems (JIFS)*, vol. 34, no. 5, pp. 2991-3001, May 2018.
8. Semantically-informed distance and similarity measures for paraphrase plagiarism identification. Miguel Ángel Álvarez Carmona, Marc Franco Salvador, Manuel Montes-Y-Gómez, Paolo Rosso, Luis Villaseñor-Pineda and **Esau Villatoro-Tello**. In *Journal of Intelligent & Fuzzy Systems (JIFS)*, vol. 34, no. 5, pp. 2983-2990, 2018.
9. Paraphrase Plagiarism Identification with Character-level Features. Fernando Sánchez-Vega, **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, Paolo Rosso, Efstathios Stamatatos, Luis Villaseñor-Pineda. (Online first version) *Pattern Analysis and Applications (PAA)*. Dic 2017.
10. Retrieving and Classifying Instances of Source Code Plagiarism. Debasis Ganguly, Gareth J.F. Jones, Aarón Ramírez-de-la-Cruz, Gabriela Ramírez-De-La-Rosa and **Esau Villatoro Tello**. (Online first version) *Information Retrieval Journal*, Vol. 21, Number 1, pp. 1-23. September 2017.
11. Early Detection of Deception and Aggressiveness using Profile-Based Representations. Hugo Jair Escalante, **Esau Villatoro-Tello**, Sara E. Garza, A. Pastor López-Monroy, Manuel Montes-y-Gómez and Luis Villaseñor-Pineda. *Expert Systems with Applications*, Vol. 89, pp. 99-111. 2017.

12. Determining and characterizing the reused Text for Plagiarism Detection. Fernando Sánchez-Vega, **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, Paolo Rosso. In *Expert Systems with Applications*, Volume 40, Issue 5, April 2013.
13. Document Ranking Refinement Using a Markov Random Field Model. **Esau Villatoro**, Antonio Juárez, Manuel Montes, Luis Villaseñor, L. Enrique Sucar. Special Issue on “*Statistical Learning of Natural Language Structured Input and Output*” of the *Natural Language Engineering Journal* 2011, Vol. 18(02), pp. 155-185.
14. Using Word Sequences for Text Summarization. **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez. In the 9<sup>th</sup> International Conference on Text, Speech and Dialogue (TSD 2006). Brno, Czech Republic. *Lecture Notes in Artificial Intelligence*, Vol. 4188, pp. 293-300, Springer 2006.

***Other Refereed Journal Publications (mainly in Spanish)***

15. Análisis de representaciones vectoriales en bitácoras de mantenimiento en la industria: Hacia un sistema de Recuperación de Información. Jesús Roberto Enrique León Carmona, Samuel González-López, **Esau Villatoro-Tello**, and Jesús Miguel García-Gorrostieta. In *Journal of Research in Computing Science*, vol 150, No. 5, pp. 231-239. Mexico City, May 2021.
16. Análisis y seguimiento de tópicos en las conferencias matutinas del presidente de México. Luis Armando Arias Romero, Gabriela Ramírez-de-la-Rosa, and **Esau Villatoro-Tello**. In *Journal of Research in Computing Science*, vol 150, No. 5, pp. 331-345. Mexico City, May 2021.
17. Inferring Highly-dense Representations for Clustering Broadcast Media Content. **Esau Villatoro-Tello**, Shantipriya Parida, Petr Motlicek, Ondřej Bojar. In *Prague Bulletin of Mathematical Linguistics (PBML) Computational Linguistics Journal*, vol. 115, pp. 31-50. October 2020.
18. Author Profiling in Social Media with Multimodal Information. Miguel Á. Álvarez-Carmona, **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, and Luis Villaseñor Pineda. In *Journal of Computación y Sistemas (CyS)*, vol. 24, no 3., pp. 1289–1304. April 2020.
19. Identificando signos de anorexia y depresión en usuarios de redes sociales. Alejandro Rosales-Martínez, Pablo Sotres, Griselda Velázquez-Villar, **Esau Villatoro-Tello** and Gabriela Ramírez-De-La-Rosa. In *Journal of Research in Computing Science*. Vol 147(7), pp. 189-202. June 2018.
20. Sistema para la generación personalizada de resúmenes a partir de múltiples documentos. Orlando Hernández-Hernández, **Esau Villatoro-Tello** and Christian Lemaître León. In *Journal of Research in Computing Science*, Vol. 147(6), pp. 85-98. June 2018.
21. Predicción de la generación de residuos sólidos urbanos en la Ciudad de México. Ester Calderón-Casanova, Mariana López, Patricia Galán, **Esau Villatoro-Tello**, Raúl R. García-Aguilar and Brenda García-Parra. In *Journal of Research in Computing Science*, Vol. 145(5), pp. 65-77. June 2018.
22. Un método para el análisis de sentimientos bajo un enfoque supervisado usando recursos léxicos. Antonio Hernández-Ambrocio, Gabriela Ramírez-De-La-Rosa and **Esau Villatoro-Tello**. In *Journal of Research in Computing Science*, Vol. 147(6), pp. 221-233. June 2018.
23. WSDL Information Selection for Improving Web Service Classification. Christian Sánchez-Sánchez, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa, Héctor Jiménez Salazar, David Pinto-Avendaño. In *Journal of Research in Computing Science*, Vol. 144, pp. 83-96. Nov 2017.

24. Resúmenes de múltiples documentos guiados por consulta empleando representaciones distribucionales. Leticia Luna-Tlatelpa, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa and Carlos Rivero-Moreno. In *Journal of Research in Computing Science*, Vol. 134, pp. 127-139. May 2017.
25. Generación y enriquecimiento automático de recursos léxicos para el análisis de sentimientos. Gerardo Real-Flores, Betzabet García-Mendoza, Ester Calderón-Casanova, Gabriela Ramírez-De-La-Rosa and **Esau Villatoro-Tello**. In *Journal of Research in Computing Science*, Vol. 134, pp. 163-176. May 2017.
26. Importancia del lenguaje coloquial y de los símbolos de puntuación en el perfilado de autores. Diana M. Sepúlveda-Barrera, Daniel Martínez-Espino, **Esau Villatoro-Tello** and Gabriela Ramírez-De-La-Rosa. In *Journal of Research in Computing Science* Vol. 115, pp. 43-56. May 2016.
27. Aplicación web para identificar personalidad, género y edad de usuarios en Twitter. Janet V. Hernández-García, Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**, Héctor Jiménez-Salazar and Verónica Reyes-Meza. In *Journal of Research in Computing Science* Vol. 115, pp. 93-106. May 2016.
28. Agrupamiento de textos cortos en dominios cruzados. Alba Núñez-Reyes, Erick Monroy-Cuevas, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa and Christian Sánchez-Sánchez. In *Journal of Research in Computing Science* Vol. 115, pp. 133-145. May 2016.
29. Exploración sobre la construcción automática de un tesoro a partir de un documento. Aarón Ramírez-De-La-Cruz, Héctor Jiménez-Salazar, **Esau Villatoro-Tello** and Gabriela Ramírez-De-La-Rosa. In *Journal of Research in Computing Science* Vol. 115, pp. 107-116. May 2016.
30. Diseño e implementación de Applets como material didáctico de apoyo para cursos de estructuras de datos. Betzabet García-Mendoza, Pablo A. Ruíz-Mendoza, Gerardo Real-Flores, Carlos R. Jaimez-González and **Esau Villatoro-Tello**. In *Pistas Educativas* Vol. 112, pp. 533-553. November 2015.
31. Plataforma web para la identificación y el análisis de eventos en Twitter. Antonio Juárez-González, Griselda Velázquez-Villar, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa. In *Journal of Research in Computing Science* Vol. 97, pp.69-78. October 2015.
32. Sistema Automático para la Clasificación de la Opinión Pública Generada en Twitter. José Ramon Gálvez-Perez, Bárbara Gómez-Torrero, Raúl Ramírez-Chávez, Kathia Sánchez Sandoval, Vicente Castellanos-Cerda, Roberto García-Madrid, Héctor Jiménez-Salazar and **Esau Villatoro-Tello**. In *Journal of Research in Computing Science* Vol. 95, pp. 23-36. May 2015.
33. Sistema Multiagente Integrador de Bibliotecas Digitales. Christian Sánchez-Sánchez, Héctor Jiménez Salazar, **Esau Villatoro-Tello**, Carlos Rodriguez-Lucatero and Gabriela Ramírez-De-La-Rosa. In *Journal of Research in Computing Science* Vol. 93, pp. 45-56. May 2015.
34. Sistema de Recomendación de Música basado en Aprendizaje Semi-supervisado. José Roberto Alvarado-García, Janet Viridiana Hernández-García, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa and Christian Sánchez-Sánchez. In *Journal of Research in Computing Science*. Vol 94, pp. 97-109. May 2015.
35. Pertinence of Lexical and Structural Features for Plagiarism Detection in Source Code. Aarón Ramírez-de-la-Cruz, Gabriela Ramírez-de-la-Rosa, Christian Sánchez-Sánchez, Héctor Jiménez-Salazar and **Esau Villatoro-Tello**. In *Journal of Research in Computing Science*. Vol. 85, pp. 21-31, Nov. 2014.

36. Detectando la Prioridad de Contenidos Generados en Twitter por medio de  $n$ -gramas de Palabras. Gerardo García-Rodríguez, Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**, Héctor Jiménez-Salazar. In *Journal of Research in Computing Science*, Vol. 74, pp. 21-32. May 2014
37. Herramienta de apoyo en la Detección de Reutilización de Código Fuente. Raymundo Picazo-Alvarez, **Esau Villatoro-Tello**, Wulfrano A. Luna-Ramírez and Carlos R. Jaimez-González. In *Journal of Research in Computing Science*, Vol. 73, pp. 45-57. May 2014
38. Laboratorio en Línea para el Procesamiento Automático de Documentos. Julio C. Torres López, Christian Sánchez-Sánchez and **Esau Villatoro Tello**. In *Journal of Research in Computing Science*, Vol. 72, pp. 23-36. May 2014
39. Clasificación de Servicios Web Empleando Estrategias Supervisadas. Victor O. Díaz-Torres, **Esau Villatoro-Tello**, Christian Sánchez-Sánchez and Héctor Jiménez-Salazar. In *Journal of Research in Computing Science*, Vol 62, pp. 109-118. April 2013
40. Análisis de Polaridad en Twitter. Gilberto Leon-Martagón, **Esau Villatoro-Tello**, Héctor Jiménez-Salazar y Christian Sánchez-Sánchez. In *Journal of Research in Computing Science*, Vol 62, pp. 69-78. April 2013
41. A Probabilistic Method for Ranking Refinement in Geographic Information Retrieval. **Esau Villatoro-Tello**, R. Omar Chavéz-García, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, L. Enrique Sucar. *Procesamiento de Lenguaje Natural*, No. 44, pp. 123-130, April 2010.
42. On the Use of Dynamic Information for Speaker Identification. Rosa M. Ortega-Mendoza, **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez, and Eduardo Morales. *Advances in Computer Science and Engineering, Research in Computing Science*, Vol. 27, pp. 161-166, April 2007.

***Refereed Conference Publications (Indexed by the Conference Proceedings Citation Index)***

43. TokenVerse: Unifying Speech and NLP Tasks via Transducer-based ASR. Shashi Kumar, Srikanth Madikeri, Juan Zuluaga-Gomez, Iuliia Nigmatulina, **Esau Villatoro-Tello**, Sergio Burdisso, Petr Motlicek, Karthik Pandia, and Aravind Ganapathiraju. *Accepted to EMNLP 2024*
44. Fast Streaming Transducer ASR Prototyping via Knowledge Distillation with Whispe. Iuliia Thorbecke, Juan Zuluaga-Gomez, **Esau Villatoro-Tello**, Shashi Kumar, Pradeep Rangappa, Sergio Burdisso, Petr Motlicek, Karthik Pandia, Aravind Ganapathiraju. *Accepted to EMNLP 2024 Findings*.
45. Mapping the Media Landscape: Predicting Factual Reporting and Political Bias Through Web Interactions. Dairazalia Sanchez-Cortes, Sergio Burdisso, **Esau Villatoro-Tello** and Petr Motlicek. (Accepted) In *Proceedings of the 15<sup>th</sup> Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF 2024 Conference and Labs of the Evaluation Forum*. September 2024.
46. Temporal fine-tuning for early risk detection. Horacio Thompson, **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, and Marcelo Errecalde. In *Memorias De Las JAIIO*, 10(1), 137-149 2024.
47. DAIC-WOZ: On the Validity of Using the Therapist’s prompts in Automatic Depression Detection from Clinical Interviews. Sergio Burdisso, Ernesto A. Reyes-Ramírez, **Esau Villatoro-Tello**, Fernando Sánchez-Vega, Adrian P. López-Monroy, Petr Motlicek. In *Proceedings of the 6<sup>th</sup>*



*Clinical Natural Language Processing Workshop @ NAACL 2024, Association for Computational Linguistics*. June 2024.

48. Reliability Estimation of News Media Sources: Birds of a Feather Flock Together. Sergio Burdisso, Dairazalia Sánchez-Cortés, **Esau Villatoro-Tello**, Petr Motlicek. In *NAACL 2024 - Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. June 2024
49. Probability-Aware Word-Confusion-Network-To-Text Alignment Approach for Intent Classification. **Esau Villatoro-Tello**, Srikanth Madikeri, Bidisha Sharma, Driss Khalil, Shashi Kumar, Iuliia Nigmatulina, Petr Motlicek & Aravind Ganapathiraju. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. April 2024
50. Multitask Speech Recognition and Speaker Change Detection for Unknown Number of Speakers. Shashi Kumar, Srikanth Madikeri, Iuliia Nigmatulina, **Esau Villatoro-Tello**, Petr Motlicek, Karthik Pandia, S. Pavankumar Dubagunta & Aravind Ganapathiraju. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. April 2024
51. Node-weighted Graph Convolutional Network for Depression Detection in Transcribed Clinical Interviews. Sergio Burdisso, **Esau Villatoro-Tello**, Srikanth Madikeri & Petr Motlicek. In *Proceedings of Interspeech 2023*. August 2023.
52. Implementing contextual biasing in GPU decoder for online ASR. Iuliia Nigmatulina, Srikanth Madikeri, **Esau Villatoro-Tello**, Petr Motlicek, Juan Zuluaga-Gomez, S Karthik Pandia D & Aravind Ganapathiraju. In *Proceedings of Interspeech 2023*. August 2023.
53. Effectiveness of Text, Acoustic, and Lattice-based representations in Spoken Language Understanding tasks. **Esau Villatoro-Tello**, Srikanth Madikeri, Juan Zuluaga-Gomez, Bidisha Sharma, Seyyed Saeed Sarfjoo, Iuliia Nigmatulina, Petr Motlicek, Alexei V. Ivanov & Aravind Ganapathiraju. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. May 2023.
54. IDIAPers @ Causal News Corpus 2022: Extracting Cause-Effect-Signal Triplets via Pre-trained Autoregressive Language Model. Martin Fajcik, Muskaan Singh, Juan Zuluaga-Gomez, **Esau Villatoro-Tello**, Sergio Burdisso, Petr Motlicek & Pavel Smrz. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*. December 2022.
55. IDIAPers @ Causal News Corpus 2022: Efficient Causal Relation Identification Through a Prompt-based Few-shot Approach. Sergio Burdisso, Juan Zuluaga-Gomez, **Esau Villatoro-Tello**, Martin Fajcik, Muskaan Singh, Pavel Smrz & Petr Motlicek. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE @ EMNLP 2022)*. December 2022.
56. Expanded Lattice Embeddings for Spoken Document Retrieval on Informal Meetings. **Esau Villatoro-Tello**, Srikanth Madikeri, Petr Motlicek, Aravind Ganapathiraju, Alexei V. Ivanov. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. July 2022
57. BertOdia: BERT pre-training for low resource Odia language. Shantipriya Parida, Satya Prakash Biswal, Biranchi Narayan Nayak, Mael Fabien, **Esau Villatoro-Tello** and Petr Motlicek. In *Proceedings of the Biologically Inspired Techniques in Many-Criteria Decision Making (BITMDM-2021)*. December 2021.
58. Automatic Dialect Detection for Low Resource Santali Language. Sunil Kumar Sahoo, Brojo Kishore Mishra, Shantipriya Parida, Satya Ranjan Dash, Jatindra Nath Besra and **Esau Villatoro-**

- Tello**. In *Proceeding of International Conference on Information Technology (OCIT)*, December 2021.
59. Approximating the Mental Lexicon from Clinical Interviews as a Support Tool for Depression Detection. **Esau Villatoro-Tello**, Gabriela Ramírez-de-la-Rosa, Daniel Gática-Pérez, Mathew Magimai.-Doss, and Héctor Jiménez-Salzar. In *Proceedings of the 23rd ACM International Conference on Multimodal Interaction (ICMI'21)*. Montreal, Canada. October 2021
  60. Late Fusion of the Available Lexicon and Raw Waveform-based Acoustic Modeling for Depression and Dementia Recognition. **Esau Villatoro-Tello**, S. Pavankumar Dubagunta, Julian Fritsch, Gabriela Ramírez-de-la-Rosa, Petr Motlicek, and Mathew Magimai.-Doss. In *Proceedings of the Interspeech 2021*. Brno, Czechia. September 2021.
  61. Open Machine Translation for Low Resource South American Languages (AmericasNLP 2021 Shared Task Contribution). Shantipriya Parida, Subhadarshi Panda, Amulya Dash, **Esau Villatoro-Tello**, A. Seza Doğruöz, Rosa M. Ortega-Mendoza, Amadeo Hernández, Yashvardhan Sharma, and Petr Motlicek. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP) at NAACL 2021*, pp. 218-223. Mexico City, June 2021.
  62. BertAA: BERT fine-tuning for Authorship Attribution. Maël Fabien, **Esau Villatoro-Tello**, Petr Motlicek, and Shantipriya Parida. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON 2020)*, pp. 127-137. India. December 2020.
  63. Detection of Similar Languages and Dialects Using Deep Supervised Autoencoders. Shantipriya Parida, **Esau Villatoro-Tello**, Sajit Kumar, Maël Fabien, and Petr Motlicek. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON 2020)*, pp. 362-367. India. December 2020.
  64. Overview of the Multimedia Information Processing for Personality & Social Networks Analysis Contest. Gabriela Ramírez, **Esau Villatoro**, Bogdan Ionescu, Hugo Jair Escalante, Sergio Escalera, Martha Larson, Henning Müller, and Isabelle Guyon. In *Pattern Recognition and Information Forensics, Lecture Notes in Computer Science, LNCS Vol. 11188*, pp 127-139. Beijing, China. August 2018.
  65. A compact representation for cross-domain short text clustering. Alba Núñez-Reyes, **Esau Villatoro-Tello**, Gabriela Ramírez-De-La-Rosa and Christian Sánchez-Sánchez. In *Advances in Artificial Intelligence, Lecture Notes in Computer Science*, vol 10061, pp. 16-26. August 2017.
  66. Enhancing Semi-Supervised Text Classification using Document Summaries. **Esau Villatoro-Tello**, Emmanuel Anguiano, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda and Gabriela Ramírez-de-la-Rosa. In *Proceedings of the 15th Ibero-American Conference on AI, IBERAMIA 2016*. Lecture Notes in Computer Science, LNCS Vol. 10022, pp 115-126. November 2016.
  67. From dialogue corpora to dialogue systems: Generating a chatbot with teenager personality for preventing cyber-pedophilia. Ángel Callejas-Rodríguez, **Esau Villatoro-Tello**, Ivan Meza, Gabriela Ramírez-de-la-Rosa,. In *Lecture Notes in Artificial Intelligence Vol. 9924*, pp. 531-539. September 2016.
  68. On the Detection of SOurce COde Re-use. Enrique Flores, Paolo Rosso, Lidia Moreno, **Esau Villatoro-Tello**. In *Proceedings of the Forum for Information Retrieval Evaluation (FIRE 2014)*, pp. 21-30. December 2015. ACM Digital Library (ISBN: 978-1-4503-3755-7).
  69. The Role of n-grams in Firstborns Identification. Gabriela Ramírez-De-La-Rosa, Verónica Reyes-Meza, **Esau Villatoro-Tello**, Héctor Jiménez Salazar, Manuel Montes-Y-Gómez and Luis

Villaseñor-Pineda. In *Advances in Artificial Intelligence and Soft Computing, Lecture Notes in Artificial Intelligence, LNAI*, vol 9413, pp. 95-106. Springer 2015 (ISSN: 0302-9743).

70. Towards Automatic Detection of User Influence in Twitter by means of Stylistic and Behavioral Features. Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**, Héctor Jimenez-Salzar, Christian Sánchez-Sánchez. In *Lecture Notes in Artificial Intelligence, LNAI* Vol. 8856, pp. 245-256, November 2014. (ISSN: 0302-9743)
71. Sexual Predator Detection in Chats with Chained Classifiers. Hugo J. Escalante, **Esau Villatoro-Tello**, Antonio Juárez-González, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, WASSA 2013, NAACL-HLT 2013*, pp. 46-54. June 2013.
72. Concept Based Representations for Ranking in Geographic Information Retrieval. Maya Carrillo, **Esau Villatoro**, Aurelio López, Manuel Montes y Luis Villaseñor. In *7th International Conference on Natural Language Processing IceTAL-2010*, Reykjavik, Iceland, August 2010. *Lecture Notes in Artificial Intelligence*, Vol. 6233, pp. 85-96. Springer 2010.
73. Ranking Refinement via Relevance Feedback in Geographic Information Retrieval. **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, and Manuel Montes-y-Gómez. In *Mexican International Conference on Artificial Intelligence MICAI 2009*. Guanajuato, Mexico, November 2009. *Lecture Notes in Artificial Intelligence* Vol. 5845, pp. 165-176. Springer 2009.
74. Multi-Document Summarization Based on Locally Relevant Sentences. **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez, and David Pinto-Avendaño. In *Mexican International Conference on Artificial Intelligence MICAI 2009*. Guanajuato, Mexico, November 09-13, pp. 87-91, 2009, *IEEE Computer Society*.
75. Representing Context Information for Document Retrieval. Maya Carrillo, **Esau Villatoro-Tello**, Aurelio López-López, Chris Eliasmith, Manuel Montes- y-Gómez and Luis Villaseñor-Pineda. Eighth International Conference on Flexible Query Answering Systems (FQAS 2009). Roskilde, Denmark, October 2009. *Lecture Notes in Artificial Intelligence* Vol. 5822, pp. 239-250. Springer 2009.
76. A Ranking Approach based on Sample Documents for Geographic Information Retrieval. **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda. Post-proceedings of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008. *Lecture Notes in Computer Science* Vol. 5706, pp. 875-879. Springer 2009.
77. A Lexical Approach for Spanish Question Answering. Alberto Téllez, Antonio Juárez, Gustavo Hernández, Claudia Denicia, **Esau Villatoro**, Manuel Montes, and Luis Villaseñor. In *Advances in Multilingual and Multimodal Information Retrieval*. Post-proceedings of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007. *Lecture Notes in Computer Science* Vol. 5152, pp. 328-331, Springer 2008.

#### ***Abstracts and Extended Abstracts in Conference Proceedings***

78. Finding Evidence of The Sexual Predators Behavior. Ángeles López-Flores, **Esau Villatoro-Tello**, and Gabriela Ramírez-de-la-Rosa. In *LatinX in AI Research at NeurIPS 2019 (LXAI 2019)*, Vancouver Canada. December 2019.
79. Mental lexicon for personality identification in texts. Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**, and Héctor Jiménez-Salazar. In *LatinX in AI Research at NeurIPS 2019 (LXAI 2019)*, Vancouver Canada. December 2019.

80. Corpus Reacción: consumers engagement in Facebook posts. Érika S. Rosas-Quezada, Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**. In *Proceedings of the Latin American and Iberian Languages Open Corpora Forum (OpenCor 2019)*, Guanajuato, Mexico. September 2019.
81. HWxPI: A Multimodal Spanish Corpus for Personality Identification. Gabriela Ramírez-De-La-Rosa, **Esau Villatoro-Tello**, Héctor Jiménez Salzar. In *Proceedings of the Latin American and Iberian Languages Open Corpora Forum (OpenCor 2018)*, Canela, Brazil. September 2018.
82. Comparación de Algoritmos de Aprendizaje para Identificación del Usuario a través de la Voz. **Esau Villatoro**, Luis Villaseñor, y Manuel Montes. *Conferencia Latinoamericana de Interacción Humano-Computadora CLIHC-05*. Cuernavaca, México, October 2005.

### *Non-Refereed Conference Publications*

83. Enhancing Multi-modal Classification of Violent Events using Image Captioning. Daniel Vallejo-Aldana, Adrián Pastor López-Monroy, **Esau Villatoro-Tello**. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023)*. September 2023.
84. The Winning Approach for the Recommendation Systems Shared Task@REST\_MEX 2022. Cipriano Callejas-Hernández, Erika Rivadeneira-Pérez, Fernando Sánchez-Vega, Adrián Pastor López-Monroy & **Esau Villatoro-Tello**. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. September 2022.
85. Leveraging Events Sub-Categories for Violent-Events Detection in Social Media. Daniel Vallejo-Aldana, Adrián Pastor López-Monroy & **Esau Villatoro-Tello**. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. September 2022.
86. Classifying the Social Media Author Profile Through a Multimodal Representation. M. Angel Álvarez-Carmona, **Esau Villatoro-Tello**, Luis Villaseñor-Pineda & Manuel Montes-y-Gómez. In: Dash, S.R., Lenka, M.R., Li, K.C., Villatoro-Tello, E. (eds) *Intelligent Technologies: Concepts, Applications, and Future Directions. Studies in Computational Intelligence*, vol 1028. Springer, Singapore. May 2022.
87. Traducción Automática de Comandos en Náhuatl para Controlar los Movimientos de un Robot Móvil. Hernández-Hernández, A., Ortega-Mendoza, R. M., **Villatoro-Tello, E.**, Camacho-Bello, C. J., & Ramírez-de-la-Rosa, G. In *Proceedings of the Segundo Congreso Internacional de Procesamiento de Lenguaje Natural para las Lenguas Indígenas*, Online, Noviembre 2021, Morelia, Michoacán.
88. Idiap and UAM Participation at MEX-A3T Evaluation Campaign. **Esau Villatoro-Tello**, Gabriela Ramírez-de-la-Rosa, Sajit Kumar, Shantipriya Parida, Petr Motliceck. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, vol. 2664, pp. 252-257. Malaga, Spain. September 2020.
89. Idiap & UAM participation at GermEval 2020: Classification and Regression of Cognitive and Motivational Style from Text. Esau Villatoro-Tello, Shantipriya Parida, Sajit Kumar, Petr Motliceck, Qingran Zhan. In *Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*, Zurich, Switzerland, June 23-25, 2020.
90. Idiap Submission to Swiss-German Language Detection Shared Task. Shantipriya Parida, Esau Villatoro-Tello, Sajit Kumar, Petr Motliceck, Qingran Zhan. In *Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*, Zurich, Switzerland, June 23-25, 2020.

91. Attribute Selection Techniques for Classification of Aggressive Tweets. LyR-UAMC participation at MexA3T 2019 Task. Gabriela Ramírez-de-la-Rosa, **Esau Villatoro-Tello**, and Héctor Jiménez-Salazar. In *Working Notes of the Iberian Languages Evaluation Forum (IberLEF) 2019*, vol. 2421, pp. 515--519. September 2019.
92. UAM's participation at CLEF eRisk 2017 task: Towards modelling depressed bloggers. **Esau Villatoro-Tello**, Gabriela Ramírez-de-la-Rosa and Héctor Jiménez-Salazar. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation forum*, CEUR Workshop proceedings Vol. 1866, Dublin, Ireland. September 2017.
93. Profile-based Approach for Age and Gender Identification. Notebook for PAN at CLEF 2016. Ma. José Garciaarena Ucelay, Ma. Paula Villegas, Dario G. Funez, Leticia C. Cagnina, Marcelo L. Errecalde, Gabriela Ramírez-de-la-Rosa and **Esau Villatoro-Tello**. In *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum*, CEUR Workshop proceedings Vol. 1609, Évora, Portugal, pp. 864-873. September 2016
94. PAN@FIRE 2015: Overview of CL-SOCO Track on the Detection of Cross-Language SOURCE COde Re-use. Enrique Flores, Paolo Rosso, Lidia Moreno, **Esau Villatoro-Tello**, Rosa Alcover and Vicente Chirivella. In *Proceedings of the Seventh Forum for Information Retrieval Evaluation (FIRE 2015)*, CEUR Workshop proceedings Vol. 1587, Gandhinagar, India, 4-6 December 2015.
95. PAN@FIRE: Overview of SOCO Track on the Detection of SOURCE COde Re-use. Enrique Flores, Paolo Rosso, Lidia Moreno, **Esau Villatoro-Tello**. In *Working Notes for the FIRE 2014 Forum for Information Retrieval Evaluation*, Bangalore, India. December 2014.
96. UAMCLyR at RepLab 2014: Author Profiling Task. Notebook for RepLab 2014 at CLEF 2014. **Esau Villatoro-Tello**, Gabriela Ramírez-de-la-Rosa, Christian Sánchez-Sánchez, Héctor Jimenez-Salzar, Wulfrano A. Luna-Ramírez, and Carlos Rodríguez-Lucatero. In *Proceedings of the Fifth International Conference on the CLEF Initiative*, Sheffield, UK. September 2014.
97. Twitter como Referente de la Opinión Pública. Ramón Gálvez Perez, Bárbara Gómez Torrero, Raúl Ramírez Chávez, Kathia Sánchez Sandoval, Vicente Castellanos Cerda, Roberto García Madrid, Héctor Jiménez Salazar, **Esau Villatoro-Tello**. In *World Association for Public Opinion Research WAPOR 2014* Santiago, Santiago de Chile. June 2014
98. UAMCLyR at RepLab 2013: Profiling Task. **Esau Villatoro-Tello**, Carlos Rodríguez-Lucatero, Christian Sánchez-Sánchez, and A. Pastor López-Monroy. In *Working Notes of the CLEF 2013 Conference and Labs of the Evaluation Forum, Valencia*, Spain. September 2013.
99. INAOE's participation at PAN'13: Author Profiling task. A. Pastor López-Monroy, Manuel Montes-y-Gómez, Hugo J. Escalante, Luis Villaseñor-Pineda and **Esau Villatoro-Tello**. In *Working Notes of the CLEF 2013 Conference and Labs of the Evaluation Forum*, Valencia, Spain. September 2013.
100. A two-step Approach for Effective Detection of Misbehaving Users in Chats. **Esau Villatoro-Tello**, Antonio Juárez-Gonzalez, Hugo J. Escalante, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda. In *Working Notes of the CLEF 2012 Conference and Labs of the Evaluation Forum*, Rome, Italy. September 2012.
101. UAM at INEX 2012 Relevance Feedback Track: Using a Probabilistic Method for Ranking Refinement. **Esau Villatoro-Tello**, Christian Sánchez-Sánchez, Héctor Jiménez-Salazar, Wulfrano A. Luna-Ramírez, Carlos Rodríguez-Lucatero. In *Working Notes of the CLEF 2012 Conference and Labs of the Evaluation Forum*, Rome, Italy. September 2012

102. INAOE at DEFT 2011: using a Plagiarism Detection Method for Pairing Abstracts-Scientific Papers. Fernando Sánchez-Vega, **Esau Villatoro-Tello**, Antonio Juárez-González, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez, and Luis Meneses-Lerín. In *Proceedings of the Seventh DEFT Workshop, DEFT2011*. Montpellier, France. July 2011.
103. INAOE at GeoCLEF 2008: A Ranking Approach based on Sample Documents. **Esau Villatoro-Tello**, Manuel Montes-y-Gómez, and Luis Villaseñor-Pineda. *Working Notes of CLEF 2008*. Aarhus, Denmark. September 2008.
104. INAOE's Participation at QA@CLEF2007. Alberto Téllez, Antonio Juárez, Claudia Denicia, Gustavo Hernández, **Esau Villatoro**, Manuel Montes, and Luis Villaseñor. *Working Notes of CLEF 2007*. 6 pags. Budapest, Hungary. September 2007
105. A Trainable Centroid-Based Document Summarizer. **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, and Manuel Montes-y-Gómez. *Workshop of Natural Language Understanding and Intelligent Access to Textual Information*. Fourth Mexican International Conference on Artificial Intelligence MICAI 2005, Monterrey, Mexico, November 2005.

#### ***Other publications (Science divulgation and opinion – non-referred)***

106. Identificación de Rasgos de Personalidad a través del lenguaje escrito. Gabriela Ramírez de la Rosa, and **Esau Villatoro Tello**. In *Panorama de las Tecnologías del Lenguaje en México*, 2<sup>nd</sup> Edición. pp.41-73. Academia Mexicana de Computación - CONACyT, México. 2019. (ISBN:978-607-97357-4-6)
107. Lingüística Computacional (Computational Linguistics). Luis Villaseñor Pineda, Noé Alejandro Castro Sánchez, Alexander Gelbukh, Yasmín Hernández Pérez, Héctor Jiménez Salazar, Manuel Montes y Gómez, David Pinto Avendaño, Fernando Sánchez , Grigori Sidorov, Gerardo Sierra Martínez, and **Esau Villatoro Tello**. In *“La Computación en México por especialidades académicas, Luis Alberto Pineda Cortés (Ed)”*, 2017.
108. Cultura Algorítmica y Tecnologías del Lenguaje Humano. Jesús Octavio Elizondo-Martínez and **Esau Villatoro-Tello**. In *Cultura Visual y Sistemas de Significación. Dando sentido a los algoritmos, los medios y la creatividad en el espacio de la comunicación*. pp. 15-28, México, UAM, Unidad Cuajimalpa. November 2015. (ISBN: 978-607-28-0520-0)
109. Recuperación de Información aplicando un Ordenamiento Basado en Documentos Ejemplo. **Esau Villatoro-Tello**, Luis Villaseñor-Pineda, y Manuel Montes-y-Gómez. *Memorias del 9o Encuentro de Investigación del INAOE*. Puebla, México. November 2008.

#### ***Book editions***

110. Intelligent Technologies: Concepts, Applications, and Future Directions, Volume 2 (1st ed.). Satya Ranjan Dash, Himansu Das, Kuan-Ching Li & **Esau Villatoro Tello** (Eds.). Springer Singapore. June 2023
111. Natural Language Processing in Healthcare: A Special Focus on Low Resource Languages (1st ed.) Dash, S.R., Parida, S., **Villatoro-Tello, E.**, Acharya, B., & Bojar, O. (Eds.). CRC Press. July 2022.
112. Intelligent Technologies: Concepts, Applications, and Future Directions Volume 1 (1st ed.). Satya Ranjan Dash, Manas Ranjan Lenka, Kuan-Ching Li & **Esau Villatoro-Tello** (Eds.). Springer Singapore. May 2022.

113. MADIC. A cinco años de su creación: pasado, presente y futuro. Martínez De la Peña, Gloria Angélica , **Villatoro Tello, Esaú**, (2019). México: UAM, Unidad Cuajimalpa, División de Ciencias de la Comunicación y Diseño, 2019

## Other Research Activities

### *Organization of international scientific workshops*

- Year: 2024* NLP Summer School- Mexican NLP Summer School co-located with NAACL 2024 (<https://ampln.github.io/escuelaverano2024/>)
- Year: 2021* NLP Summer School- Mexican NLP Summer School co-located with NAACL 2021 (<https://ampln.github.io/escuelaverano2021/>)
- Year: 2018* HwxPI shared task - 2018 ICPR Multimedia Information Processing for Personality & Social Networks Analysis Challenge co-located with ChaLearn, ImageCLEF, MediaEval, IAPRTC12 (<https://chalearnlap.cvc.uab.cat/challenge/27/description/>)
- Year: 2015* CL-SOCO - Detection of Cross-lingual Source COde Re-use 2015 at FIRE 2015 (<https://pan.webis.de/fire15/pan15-web/source-code-reuse.html>)
- Year: 2014* SOCO - Detection of Source COde Re-use 2014 at FIRE 2014 (<https://pan.webis.de/fire14/pan14-web/index.html>)

### *Participation in International scientific competitions*

- Year: 2023* First place on the [DA-VINCIS](#): Detection of Aggressive and Violent Incidents from Social Media in Spanish task. Our proposed multimodal approach was able to efficiently and effectively detect violent events reported on social media among all the competitive systems evaluated during the DA-VINCIS 2023 edition.
- Year: 2022* First place on the [Rest-Mex 2022](#): Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts. Overall, our proposed approach, working in collaboration with the CIMAT Research Institute from Mexico, was able to obtain the best performance among competitors during the Rest-Mex 2022 edition.
- Year: 2022* Outstanding result in the [CASE @ EMNLP 2022](#) shared tasks: Challenges and Applications of Automated Extraction of Socio-political Events from Text. Our team developed a cause-effect identification system based on a Prompt-based Few-shot Approach.
- Year: 2021* Outstanding results in US. IARPA(Material). Overall the SARAL team (Idiap and ISI) obtained excellent results across multiple languages in the US. [IARPA \(Material\)](#) project. During my collaboration with the team, the SARAL group (Idiap and ISI) ranked 1st in speech and 3rd in text for the Kazakh topical broadcast retrieval.

- Year: 2021*      Second place in the Open Machine Translation for Low resource South American Languages ([AmericasNLP@NAACL 2021](#)). Our approach obtained the second-best results for the language pairs “Spanish-Bribri”, “Spanish-Asháninka”, and “Spanish-Rarámuri”.
- Year: 2020*      First place on the Fake news detection task during the [MEX-A3T@IberLEF](#): Fake News and Aggressiveness Analysis. Our proposed system employs Supervised Autoencoders (SAE) as main learning algorithm in the text classification task.
- Year: 2020*      Second place in the [GermEval Task 1](#) - “Classification and Regression of Cognitive and Motivational Style from Text”. Our proposed method is based on Transformers architecture.
- Year: 2020*      Third place in the [Swiss German Language Detection shared task](#). We evaluated the impact of supervised autoencoders.
- Year: 2013*      Second place in the Author Profiling shared task held at [PAN@CLEF 2013](#). Our team developed a second-order representation approach for social media documents, showing excellent performance in age and gender classification problems.
- Year: 2012*      First place in the Sexual Predator Identification in online chat rooms task during the [PAN@CLEF 2012](#). Our team developed a two-step approach to detect predators in online chats. The first module detects suspicious conversations, while the second module distinguishes victim and predator.

### ***Memberships***

- Member of Mexico’s National System of Researchers SNI-CONACyT (Awarded distinction level 1) from 2020 to the present.
- Active member of the Mexican Association of Natural Language Processing (AMPLN) since 2020.
- Active member of the Mexican Academy of Computer Science (AMEXCOMP) since 2015.
- Regular reviewer and member of the Programm Committee of ACL, EACL, EMNLP, COLING, NAACL, LREC, CoNLL, and IJCNLP conferences
- Regular Area Chair of the ACL Rolling Review (ACL\*) system

## **Teaching and Supervision Record**

### ***Teaching Graduate Courses***

- Social Networks Analysis in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Period: Winter 2019*
- Text Classification in Universidad Autónoma de Querétaro, Campus Aeropuerto  
*Period: Spring 2018*
- Natural Language Processing Techniques in Universidad Autónoma Metropolitana Unidad Cuajimalpa



*Period: Winter 2014, Winter 2015, Winter 2016, Winter 2017, Winter 2018*

### **Teaching Undergraduate Courses**

- Artificial Intelligence II (Machine Learning) in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Periods: Winter 2013, Autumn 2014, Autumn 2015, Winter 2015, Summer 2016*
- Introduction to Logical Thinking in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Period: Autumn 2012*
- Artificial Intelligence I in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Periods: Autumn 2012, Autumn 2016, Autumn 2018*
- Data Structures in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Periods: Summer 2017, Summer 2018*
- Fundamentals of Programming in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Periods: Autumn 2013, Winter 2014, Winter 2019, Winter 2020*
- Software Engineering in Universidad Autónoma Metropolitana Unidad Cuajimalpa  
*Periods: Summer 2013, Summer 2014, Summer 2015, Winter 2016, Autumn 2016, Winter 2017, Autumn 2017, Winter 2018*

### **List of supervisions**

- **Postdoc (1)**  
*Fernando Sánchez Vega - 2 years, at Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.*
- **PhD students (4)**
  - *Gabriela Ramírez de la Rosa (thesis public defense- 26th July, 2024), at Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.*
  - *Horacio Thomson (to be concluded in 2025), at Universidad Nacional de San Luis, Argentina.*
  - *Amadeo M. Hernández Hernández (to be concluded in 2024), Universidad Politécnica de Tulancingo, Hidalgo, Mexico.*
  - *Miguel Ángel Álvarez Carmona (concluded in March 2019), at INAOE, Puebla, Mexico.*
- **Master students (9)**
  - *Ernesto Antonio Reyes Ramírez (to be concluded in November 2024) at Centro de Investigación en Matemáticas (CIMAT), Mexico.*
  - *Mariana Valeria López Ortíz, María Ester Calderón Casanova, Patricia Galán Lara, María Elena Sánchez Vilchis, José Ramón Gálvez Pérez, Bárbara Gómez Torrero, Raúl Ignacio Ramírez Chávez, Kathia Michel Sánchez Sandoval at Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.*
- **Bachelor students (17)**
  - *Raymundo Picazo Álvarez, Julio Cesar Torres López, Víctor Oscar Díaz Torres, Griselda Velázquez Villar, Ángel Callejas Rodríguez, Janet Viridiana Hernández García, Emmanuel Ulisses González López, José Antonio Hernández Ambrocio, Orlando Hernández Hernández, Jorge Elías Vigil Santos, Alejandro Díaz Ávalos, Tonantzin*

*Angélica Siurob Palomero, Érika Sarai Rosas Quezada, Ángeles López Flores, Gerardo Uriel Ruiz Santiago, Luis Armando Arias Romero, Nathalie Reyes Zauco* at Universidad Autónoma Metropolitana Unidad Cuajimalpa, Mexico.

- **Interns (4)**
  - Alexandre Huohu, Erik Börve, Gonçalo Gomes, Marcus Henriksbø (EPFL students) at Idiap Research Institute, CH

#### Author Identifiers

- ORCID: 0000-0002-1322-0358 [LINK](#)
- SCOPUS ID: 5043373400 [LINK](#)
- Google Scholar: --- [LINK](#)