Felipe **González-Pizarro**

MSc. in Computer Science · Bs. in Computer Science and Engineering

+1 (236) 591 2455 | fpeandrees@gmail.com | http://gonzalezf.github.io/ | GonzalezF | FelipeGonzalezPizarro | @fpeandrees

Education

MSc. in Computer Science

Vancouver, Canada

The University of British Columbia

Sept. 2021 - July 2023

- Focus on Natural Language Processing, Multimodal Machine Learning, and Information Visualization. **Average grade 95%. GPA: 4.0/4.0**
- Relevant coursework: Multimodal Learning with Vision, Language, and Sound (CPSC 532S), Discourse in NLP (532G), Commonsense Reasoning in NLP (CPSC 532V), Computational Linguistics (CPSC 503), Information Visualization (CPSC 547), Topics in Human-Computer Interaction (CPSC 554)

MSc. in Computer Science

Santiago, Chile

Universidad Técnica Federico Santa María

March 2018 - Sept. 2021

- Focus on Social Computing using Natural Language Processing, Deep Learning and Data Visualization methods. **Average grade 91%. GPA: 4.0/4.0**

Bsc. in Computer Science and Engineering

Santiago, Chile

Universidad Técnica Federico Santa María

March 2012 - Feb. 2018

- Comprehensive six-year program with a strong emphasis on Software Engineering, Project Management and Information Retrieval.
- Average grade 81%, GPA: 3.7/4.0, Passed subjects: 66/66. **Best Graduated Student, Rank: 1/32**

Skills

Natural Language Processing	Building and training ML models for NLP tasks using NLTK, spaCy, scikit-learn, Gensim, and HuggingFace.
Multimodal Machine Learning	Deep learning for vision and language using PyTorch (e.g., VAEs, RNNs, GANs, LSTMs, CNNs, Difussion models).
Information Visualization	Interactive visualizations with D3.js, Javascript, JQuery, Plotly, Seaborn, and Matplotlib.
Human Computer Interaction	Conducting user studies, prototyping, and designing effective interfaces. Statistical analysis with Python/R.
Programming and Tools	Python (Pandas, Numpy, Tensorflow, Django, Flask, FastAPI), Javascript (D3.js, JQuery), C/C++, Git, Unix
Languages	English (IELTS Overall Band Score: 7.0), Spanish (Native), Italian (Elementary proficiency)

Selected Work Experience

Graduate Teaching assistant

Vancouver, Canada

The University of British Columbia

Sept. 2021 - Present

- Courses: Topics in Computer Science - Natural Language Processing (CPSC 436N); Advanced Methods for Human Computer Interaction (CPSC 444); Applied Machine Learning (CPSC 330); Basic Algorithms and Data Structures (CPSC 221).
- Graded assignments and exams, led discussion sections, held office hours, prepared course materials, and provided constructive feedback to students.

Visiting Scholar

Saarbrücken, Germany

Max Planck Institute for Informatics

June 2021 - Aug. 2021

- Investigated the potential of large vision pre-trained models based on Contrastive Learning (e.g., CLIP) to aid in the detection of hateful imagery.
- Developed a methodology utilizing Google's Perspective API and manual annotations to identify Antisemitic/Islamophobic textual phrases, leveraging the CLIP model to identify hateful imagery based on these phrases outperforming baseline hateful detection methods.
- Developed a methodology utilizing Google's Perspective API and manual annotations to identify Antisemitic/Islamophobic textual phrases in a large dataset of 66 million social media posts. Leveraged the CLIP model to enhance the detection of hateful imagery based on these phrases, surpassing the performance of baseline hateful detection methods.
- Published 1 paper on a highly-ranked international computer science conference (ICWSM'23, Acceptance rate: 20%).

Visiting Researcher

Halifax, Canada

Dalhousie University

Jan. 2020 - Dec. 2020

- Developed TopicVisExplorer, a web-based interactive visualization tool that enables users to refine and compare topic models from large-scale datasets.
- Proposed novel topic refinement operations and a topic similarity metric to support users in the interpretation, improvement, and comparison of models.
- Designed and conducted a user study with over 40 non-NLP experts to demonstrate the effectiveness of TopicVisExplorer in facilitating human interpretation and comparison of LDA-generated topics. Analyzed the collected data to validate the findings.

Research Assistant

Santiago, Chile

University of Washington - Universidad Técnica Federico Santa María

March 2018 - Dec. 2019

- Proposed a novel methodology for inter-language comparison of social media text that offers an alternative method to conduct studies on data privacy perspectives across speakers of different languages and provide a roadmap for future cross-cultural research.
- Collected and analyzed unstructured textual social media data related to information privacy. Inter-language differences found on privacy-related views expand current knowledge of information privacy perspectives.
- Published 3 papers on highly-ranked international computer science venues (WWW'19, CSCW'19, ECSCW'22) .

Selected Peer-Reviewed Publications

Understanding and Detecting Hateful Content using Contrastive Learning

In International Conference on Web and Social Media (ICWSM'23)

- Explored the efficacy of OpenAI's CLIP, a large vision-language pre-trained model, in effectively detecting Antisemitic/Islamophobic imagery, showcasing its remarkable performance with an accuracy of 81% and offering new possibilities for hate speech detection.
- Implemented an innovative approach combining manual annotations, Google's Perspective API, and the CLIP model to accurately identify Anti-semitic/Islamophobic textual phrases and detect associated hateful imagery, outperforming baseline methods.

Diversity-Aware Coherence Loss for Improving Neural Topic Models

Association for Computational Linguistics (ACL'23)

- Collaborated in proposing a novel diversity-aware coherence loss function, enhancing the coherence and diversity of neural topic models concurrently.
- Integrated our innovative approach into OCTIS, a Python library for optimizing and comparing topic models, expanding its functionalities, and empowering researchers with advanced techniques to enhance the coherence and diversity of topic models.

TopicVisExplorer: An interactive visualization tool to refine and compare topic models

In ACM Transactions on Interactive Intelligent Systems (Submitted)

- Developed TopicVisExplorer, a web-based interactive visualization tool, to refine and compare topics generated by LDA. Implemented novel features, such as topic refinement operations and a new topic similarity metric, enabling users to refine (i.e., split, merge) and compare topics effectively.
- Conducted and analyzed a user study with over 40 participants, demonstrating TopicVisExplorer's support in identifying topics for refinement and reducing erroneous matches in two-corpora comparisons, highlighting its practical utility in topic modeling tasks.

Inequalities in Computational Thinking among Incoming Students in a STEM Chilean University

IEEE Latin America Transactions Journal

- Conducted a cross-sectional survey among more than 500 incoming students of a Chilean STEM university to assess computational thinking skills.
- Analyzed survey data from over 500 respondents using statistical analysis techniques in R, revealing significant inequalities in computational thinking skills based on gender, type of school, and prior programming knowledge.

Regional Differences in Information Privacy Concerns After the Facebook-Cambridge Analytica Data Scandal

The Journal of Collaborative Computing and Work Practices (JCSCW, ECSCW'22)

- Collected, preprocessed, and analyzed a large dataset consisting of 7.4M English tweets and 480K Spanish tweets associated with a data privacy scandal.
- Constructed and evaluated word embeddings using various architectures (e.g., Word2Vec, FastText). Ensured fairness by removing gender bias from them.
- Conducted open coding and content analysis to develop a comprehensive categorization based on the data. Explored the implications of our findings, expanding the current understanding of privacy concerns on a global scale.

Information Privacy Opinions on Twitter: A Cross-Language Study

Computer Supported Cooperative Work and Social Computing (CSCW'19)

- Constructed and evaluated word embeddings to capture the semantic nuances of Spanish and English tweets, providing insights into information privacy perspectives across Twitter users during a data privacy scandal. Employed open coding methods to categorize and characterize the data.

Global Reactions to the Cambridge Analytica Scandal: A Cross-Language Social Media Study

In Companion Proceedings of The 2019 World Wide Web Conference (WWW'19)

- Conducted a large-scale cross-language analysis of 8 million tweets in Spanish and English on the #CambridgeAnalyticaScandal. This study presents a novel methodology for exploring and comparing social media data, offering an alternative approach to traditional survey-based methods.
- Played a key role in pioneering research on the Facebook-Cambridge Analytica Scandal by publishing a groundbreaking paper, which has been widely cited 46 times, demonstrating the enduring impact and relevance of the study in the field.

Development of Computational Thinking in High School Students: A Case Study in Chile

Best paper award

International Conference of the Chilean Computer Science Society IEEE (SCCC'18)

- Designed and executed a case study in Chile, conducting a 15-week voluntary C++ workshop for 99 high school students. The workshop aimed to investigate the development of computational thinking skills.
- Evaluated the impact of the C++ workshop on enhancing computational thinking abilities by conducting statistical analyses using Python and R, employing various tests such as ANOVA and t-tests. The findings were documented in a research paper, which was recognized as the best paper in the field.

Selected Honors & Awards

2021	**Computer Science Merit Scholarship,** The University of British Columbia (UBC). Outstanding Graduate Student applicant. CAD 20,000 for living expenses	*Canada*
2020	**Emerging Leaders in the Americas Program (ELAP),** EduCanada Research Scholarship. Funding for research in Canada at Dalhousie University (Canada). USD 7,300 for travel and living costs	*Canada*
2019	**The Cornell, Maryland, Max Planck Pre-doctoral Research School (CMMRS 2019) ,** Selected participant at CMMRS 2019, a fully-funded opportunity at Max Planck Institute (MPI-SWS) to explore cutting-edge computer science research.	*Germany*
2019	**National MSc. Grant,** National Commission for Scientific and Technological Research (CONICYT). Top 7% Applicant. USD 9500 for tuition and living expenses.	*Chile*