Felipe Russi

feliprussi@gmail.com | (+57) 30594139887 | Github | LinkedIn | Portfolio

EDUCATION

Universidad de los Andes, *Bogota, Colombia*
BSc in Systems and Computing Engineering, GPA: 3.8/4.0 *Aug 2019 - December 2024*
Universidad de los Andes, *Bogota, Colombia*
Bachelor of Mathematics, GPA: 3.8/4.0 *Aug 2020 - December 2024*
Relevant Coursework: *Machine Learning, Generative Models, Natural Language Processing, Quantum Computing, Data Structures, Design and Analysis of Algorithms, Business Intelligence, Computing Infrastructure, Database Systems, Statistics, Probability, Pattern Recognition, Numerical Analysis, Differential Geometry, Topology, Logic, Information Theory, Abstract Algebra, Measure and Integration Theory.*

GRANTS AND AWARDS

August 2019. *Generación E Scholarship, Universidad de los Andes and Ministry of National Education (Colombia).*
January 2024. *DevSavant-IntelePeer Excellence Scholarship, DevSavant and IntelePeer.*
June 2024. *Summer Undergraduate Research Fellowship (SURF), Universidad de los Andes and Cornell Tech.*
August 2024. *Vamos Pa'lante Scholarship, Universidad de los Andes.*

RESEARCH PAPERS

A Multi-Agent Framework with Diagnostic Feedback for Iterative Plain Language Summary Generation from Cochrane Medical Abstract
- Designed a multi-agent LLM system to generate Plain Language Summaries for biomedical evidence.
- Implemented specialized agents and analytic tools to improve summary clarity, factuality, and interpretability.
- Ran large-scale evaluations on 100 Cochrane abstracts using six LLMs, identifying trade-offs between readability and accuracy.
- *Submitted to TSAR @ EMNLP 2025. Accepted on September 30th 2025 (check here).*

Bridging the Gap in Health Literacy: Harnessing the Power of Large Language Models to Generate Plain Language Summaries from Biomedical Texts
- Conducted research to improve health literacy by using **LLMs** to generate simplified biomedical text summaries.
- Built and tested **linguistic analysis methods** to evaluate the generated text for simplicity and semantic fidelity.
- Developed an API to support plain language analysis and facilitate medical text translation with LLMs.
- *Submitted to CL4Health @ NAACL 2025. Accepted on March 3rd (check here).*

Guiding Multimodal Large Language Models with Blind and Low Vision Visual Questions for Proactive Visual Interpretations
- Developed a system that improved MLLM descriptive accuracy for blind users by 13.1% using historical user questions.
- Engineered a RAG pipeline that successfully anticipated user needs in 15.2% of cases where the baseline model failed.
- *Submitted to CV4A11y @ ICCV 2025,(check OpenReview here).*

CARIS: A Context-Adaptable Robot Interface System for Personalized and Scalable Human-Robot Interaction
- Designed a modular robot interface system adaptable across diverse environments.
- Focused on enhancing **user experience** by optimizing **LLM-based prompts** and intuitive **communication options.**
- Developed a **scalable human-in-the-loop interface** that supports **real-time decision-making** for behavioral modeling.

RESEARCH EXPERIENCE

Kelsenia | *Founding Software Engineer* *March 2025 – Present, Bogota, Colombia*
- Engineered a Legal AI Agent using a RAG system to deliver high-accuracy, up-to-date and verifiable answers on Colombian law.
- Built and deployed the full-stack legal platform using FastAPI, PostgreSQL/Supabase, and AWS.
- Co-designed and validated AI-powered legal tools through direct collaboration with a law firm to solve real workflow challenges.
- Developed a custom framework to automate AI-driven legal workflows, boosting lawyer productivity.

FLAGLab at Universidad de los Andes | *Research Assistant* *Aug 2022 – Present, Bogota, Colombia*
- Developed a custom metric to measure text plainness and built **ML models** to classify text as plain or professional.
- Collaborated with professionals in **medical writing** from **MSD** to define key challenges in **health literacy**.
- Implemented **readability analysis** using **NLP libraries**, enhancing user experiences through accessible text.
- Developed and **fine-tuned NLP models** for semantic classification, enhancing plain-language **generation pipelines.**
- Managed **large biomedical datasets** with tools like Pandas and NumPy, and collected data using web scraping.
- Conducted **statistical analysis** to optimize **LLM text generation** and ensure adaptability across diverse domains.

AIRLab at Cornell University | *Research Assistant* *Jun 2024 – March 2025, New York City, USA*
- Led a project to develop a web interface to support healthcare workers through adaptable robot systems.
- Developed a web interface using **Svelte** and **FastAPI** for controlling a **TIAGo base robot**, integrating cross-functional tools.
- Evolved the interface from **medical assistance** to **general-purpose control**.
- Reduced **latency by 80%** through optimized WebRTC communication and **ROS bridge** integration.
- Enhanced **vision tracking** with **YOLO** and optimized **dialogue interactions** and system prompts using **LLMs.**

- Conducted **pilot studies** and analyzed workflows to improve usability and functionality.

Universidad de los Andes | *Mathematics Research Assistant*   *Jan 2024 – Jan 2025, Bogota, Colombia*
- Contributed to a **seminar on Optimal Transport** and expanded this research area within the university.
- Studied **Mean-Field Variational Inference** and **Optimization** to understand its applications in probabilistic models.
- Explored Optimal Transport to improve the accuracy and convergence of **variational methods.**
- Applied known inference and transport methods to **generative models** and **clustering**.

Cornell Tech | *Research Intern*   *Jun 2024 – Aug 2024, New York City, USA*
- Developed a web interface using **Svelte** and **FastAPI** for controlling a **TIAGo base robot**.
- Evolved the interface from **medical assistance** to **general-purpose control**.
- Implemented **scalable architectures** ensuring **low-latency interactions** with WebRTC and rosbridge.

ML4I at Universidad de los Andes | *Machine Learning Specialist*   Jan 2024 – Jun 2024, Bogota, Colombia
- Conducted research on simulating data for financial sector risk analysis for a **Colombian stock exchange company**.
- Developed machine learning models to predict and mitigate financial risks based on simulated data.
- Collaborated with a cross-functional team to integrate ML solutions into financial decision-making processes.

Robocol at Universidad de los Andes | *Computer Vision Engineer*   *Jan 2021 – Aug 2022, Bogota, Colombia*
- Developed and optimized object detection algorithms using **YOLO** and **CNNs** for robotics competitions.
- Collaborated with multidisciplinary team members to integrate **vision systems** into robots and improve performance.
- Conducted field tests in desert environments to evaluate and optimize vision-based tasks for the rover.
- Reached final stages in the **URC 2022**.

WORK AND TEACHING EXPERIENCE

Universidad de los Andes | *Secondary Instructor*   Jan 2024 – *Jun 2024, Bogota, Colombia*
- Taught one semester of **Differential Calculus** to +30 students, with 2 classes per week.
- Led workshops and quizzes and provided office hours to reinforce understanding and evaluate progress.
- Created course materials, problem sets, and assessments to enhance learning and comprehension.

Universidad de los Andes | *Math Tutor*   Aug 2023 – *Jun 2024, Bogota, Colombia*
- Assisted students with **math subjects** including Pre-Calculus, Differential and Integral Calculus, and Linear Algebra.
- Conducted **6 hours of weekly** tutoring sessions, providing personalized support to reinforce core concepts.
- Applied **active learning strategies** to engage students and improve their problem-solving skills.

Universidad de los Andes | *Teaching Assistant*   Jan 2022 – *Dec 2022, Bogota, Colombia*
- Supported +**50 students** in the **Algorithm Design and Analysis** course through grading and review sessions.
- Clarified **algorithmic concepts** and provided individualized support.
- Guided hands-on exercises in **algorithm design** and **complexity analysis**.
- Assisted with **data structure optimization** through coding tasks.
- Designed and delivered **interactive coding workshops** to enhance collaborative learning.

PROJECTS

CARIS: A Context-Adaptable Robot Interface System for Personalized and Scalable Human-Robot Interaction
- Designed a modular robot interface system that combines **navigation**, real-time **comprehension**, and multimodal data recording to improve adaptability across diverse environments.
- Focused on enhancing **user experience** by refining interaction flows, optimizing **LLM-based prompts**, and incorporating intuitive **communication options.**
- *Planned for submission to RO-MAN 2025.*

Bridging the Gap in Health Literacy by NLP and LLMs
- Conducted research to improve health literacy by using **LLMs** to generate simplified biomedical text summaries.
- Built and tested **linguistic analysis methods** to evaluate the generated text for simplicity and semantic fidelity.
- Developed an API to support plain language analysis and facilitate medical text translation with LLMs.

Generative Models for Quasar Spectra
- Trained **Variational Autoencoders (VAEs)** and **GANs** using **TensorFlow** to generate **quasar spectra data**.
- Created synthetic quasar data with modified parameters to simulate astronomical scenarios.
- Collected an extensive quasar spectra dataset.

DevSavant Contest
- Collaborated on a team to develop a Visual Studio Code extension for project analysis using **LangChain** and **LLMs**.
- Extracted key project information and used LLMs to answer questions about **code functionality** and **modifications**.
- Enabled efficient **code analysis,** helping developers understand and improve code quality.

Mobile Application for Restaurant Discovery
- Collaborated with a team of six to develop a **dual-platform mobile app** using **Kotlin** for Android and **Flutter.**

- Designed the app to help users discover and publish restaurants near the university.
- Focused on **software architecture**, **optimization**, and **multithreading** to ensure seamless functionality.

SKILLS SUMMARY

Soft Skills: Collaborative problem solver, effective communicator, experienced in education, open-minded, adaptable to new challenges, resourceful, excels in teamwork, and maintains a positive attitude.

Programming Languages: Python, Java, R, Dart, JavaScript, SQL, HTML\CSS.

Tools/Frameworks: FastAPI, Tensorflow, Keras, Pytorch, spaCy, scikit-learn, Flutter, Docker, PostgreSQL, ROS.

Expertise: Data {Modeling, Analytics, Visualization}, ML Pipelines, NLP, Deep Learning, LLMs, Optimization, Software Architecture, HCI, HRI.

CERTIFICATIONS

NVIDIA – Building Transformer-Based Natural Language Processing Applications – *May 2024*

Coursera – DeepLearning.AI TensorFlow Developer – *Jun 2021*

Coursera – Deep Learning Specialization – *May 2021*

Coursera – Machine Learning (Stanford University) – *Jul 2020*