# Juan Nicolas Sepulveda Arias

MSc (Computer Vision) @ MBZUAI | AI Safety & Security | Theoretical Grounding

Abu Dhabi | jn.sepulveda55@gmail.com | +971 58 538 1633 | linkedin.com/in/juannicosepulveda

jnsepulveda55.github.io | github.com/JNSepulveda55 | orcid.org/0009-0000-0619-0704

## Research Interests

**AI safety & fairness for generative models: adversarial robustness (jailbreaks/red-teaming) and content authenticity (watermarking for diffusion/LLMs).**

## Education

**Mohamed Bin Zayed University of Artificial Intelligence**, *MSc in Computer Vision* Aug 2025 – May 2027

- Awarded **fully funded** scholarship by the UAE government.
- Courses include *Fairness and Safety in Generative AI*, *Machine Leaning Security*, *Deep Learning for Visual Computing* and *Mathematical Foundations of Artificial Intelligence*.

**Universidad del Rosario**, *BSc in Applied Math and Computer Science* Jan 2021 – Dec 2024

- GPA: 4.62/5.0
- Awarded **50%** scholarship being a top applicant.
- Recognition of excellence for graduating with the **top CGPA** in my batch.

## Skills

- ML: PyTorch, HuggingFace, Lightning, FastAI
- Math/Tools: MATLAB, Mathematical Proofs
- Development: GCP, Django, FastAPI, Git/Github
- Languages: English (C1), Spanish (native); Portuguese (B2).

## Experience

**Graduate Researcher** MBZUAI - Abu Dhabi, UAE Aug. 2025 - Ongoing

- Conducting research under the supervision of Professor Nils Lukas.
- Performing literature review on AI fairness metrics and state of the art content watermarking methods.

**RLHF/SFT - Python coder** Turing - Remote Sep. 2024 – Dec. 2024

- Produced avg. 10/day high-quality SFT/RLHF training examples for code-writing and problem-solving.
- Applied rubric-based evaluation to identify failure modes (hallucination, instruction-following errors, abstract reasoning).

**Software Engineering Intern**, DevSavant – Bogota, Colombia July 2023 – Jan. 2024

- Developed an AI-Powered MVP that assists elderly people in writing their autobiography with the help of LLMs.
- Developed a web application using Django and deployed the web application on a Google Cloud Platform VM for $\sim$ 10 beta testers to access and give feedback.
- Connected the web app with the OpenAI API to get responses from user-customized prompts.
- Structured the project into small tasks suited for two-week sprints.

**Tutor**, Universidad del Rosario – Bogota, Colombia Jan. 2022 – July 2024

- 2 Weekly sessions with $\sim$ 15-30 students. Solved student questions and offered guidance on course contents.
- Created problem sets and real-world case studies.
- Courses I tutored: *Intro to Data Analysis and Deep Learning* ***(graduate)****, Differential Calculus, Linear Algebra, Probability & Statistics, Algorithms & Data Structures* ***(undergraduate)***

## Projects

**Using Generative Adversarial Networks (GANs) to Detect Cyberthreats**

- Explored performance of a **GAN** in detecting and generating non-obfuscated and obfuscated binary malware samples.
- Trained on the VirusShare dataset and observed a 3% (92% to 95%) accuracy improvement on benchmarks compared to a previous **CNN** approach.
- Applied **interpretability** techniques like **hooks**, **GradCAM** and binary reverse engineering.
- Presented the project at **BSIDES Colombia 2024** cybersecurity conference.

**Traffic Signs Recognition Systems**

- **Fine-tuned** model YOLOv8 with the GTSRB dataset to detect traffic signs under different weather conditions.
- Used **data augmentation** techniques to simulate fog, night and rain conditions.
- Tested different architectural changes to the last layer and $mAP_{50-95}$ jumped from 0.39 to 0.65.

## Extracurricular

**President of the IEEE Computer Society student chapter** Feb. 2024 - Dec. 2024
*Universidad del Rosario*

- Organized Internal programming competitions $\sim$ 15-20 participants.
- Gave AI skills workshop for company executives and graduate students.
- Hackathon for high schoolers called "Code The Future" with 60 students into 15 teams.
  - Led "Code The Future" hackathon with 60 high school students (15 teams) in collaboration with UAE Embassy in Colombia. Trained participants on Gen-AI fundamentals to develop solutions for rising sea levels. Top 5 finalist teams presented to the UAE Ambassador and Dean of Engineering.