

Personal information & contact

- Nationality: Spanish, Colombian
- Location: Barcelona, Spain
- Telephone: +34 644092150
- Email: carlosgg33@gmail.com
- [linkedin.com/in/carlosggog](https://www.linkedin.com/in/carlosggog)
- github.com/carlos-gg
- [Google scholar](https://scholar.google.com/citations?user=...)
- [Personal website](#)
- [Knowledge garden](#)

Education

- PhD in Sciences - Astrophysics and Computer Vision, *Universite de Liege, Belgium*
- MSc in Astrophysics, *Universidad Complutense and U. Autónoma de Madrid, Spain*
- Specialization in Software Development, *Universidad del Magdalena, Colombia*
- BSc in Astronomy, *St. Petersburg State University, Russia*

Training and courses

- [LLMOps](#) (2024, DeepLearning.AI)
- [AI Agentic Design Patterns with AutoGen](#) (2024, DeepLearning.AI)
- [Functions, Tools and Agents with LangChain](#) (2024, DeepLearning.AI)
- [LangChain for LLM Application Development](#) (2023, DeepLearning.AI)
- [Building Systems with the ChatGPT API](#) (2023, DeepLearning.AI)
- Deployment of Machine Learning models on Azure cloud (2023, The Bridge)
- Big data analytics (2020, PRACE Advanced Training Centres)
- ESA bootcamp for Earth observation innovation and startup creation (2018, ESRIN Italy)
- [PRAIRIE Artificial Intelligence Summer School](#) (2018, INRIA France)
- NUMEDIART Deep Learning workshop (2017, Universite de Mons Belgium)
- Courses on ML and statistical inference at the Montefiore Institute (2014, University de Liege, Belgium)

Carlos Alberto Gomez Gonzalez

Lead data scientist

With over 7 years of experience as data scientist and AI engineer, I have cultivated a deep understanding of AI-driven innovation and the intricacies of leading end-to-end ML projects across diverse sectors. My expertise ranges from orchestrating impactful R&D initiatives in Space and Earth sciences to driving ML projects in dynamic industries such as retail, manufacturing, and insurance within the private sector.

Work Experience

Lead Data Scientist

2023 June – Present

[Decide4AI \(Decide Soluciones\)](#)

- Led the design and directed the development of the following projects:
 - Multi agent-based system with text-to-SQL and dynamic reporting capabilities for a SaaS workforce management company. This system is capable of generating automated visualizations and natural language explanations in response to user questions related to data stored in a PostgreSQL database. The multi-agent system was implemented with Autogen, OpenAI LLMs and validated using the AgentOps LLMOps platform.
 - Document processing system for an InsurTech Spanish startup, enhancing their legacy pipeline built on Google cloud with AutoML models. Integrated custom text classification models and LLM-based (GPT-4o) knowledge extraction techniques to reduce the cost of inference by up to 50% and to improve relevant KPIs by a factor of 4.
 - Demand forecasting system for a large international power tools manufacturer. This solution accurately projected unit sales up to a 12-month horizon, and was containerized using Docker.
 - Credit scoring system for a Chilean credit cooperative. Relevant metrics were improved up to 20% over the baseline. The system API was built using fastAPI ensuring streamlined deployment.
 - End-to-end forecasting solution, with a 3 months horizon, tailored for a Chilean hygiene products manufacturer. This solution integrated a demand elasticity model for strategic decision-making. Our solution had very good MAPE metrics from 5% to 40% (for the most difficult time series) and was containerized with Docker to facilitate its deployment.
- Developed internal tools and POCs for commercial demos focused on the usage of LLMs and multi-agent systems, for instance:
 - POC of a contract audit system integrating Computer Vision models for signature detection (fine-tuned YOLO), OCR and information retrieval via prompt engineering and RAG, using LangChain and the OpenAI API.
 - Text categorization pipeline tailored for customer service dialogues, comparing diverse methodologies, including ML classification over TF-IDF encoding, fine-tuning of BERT, zero and few-shot learning with GPT-4, and both few-shot and LoRA fine-tuning with Llama3-8B-instruct.
- Carried out research and development internal activities related to Causal ML by developing a python library for cost-sensitive uplift modeling.
- Provided support to data science teams working with external clients (InsurTech and energy sectors).
- Engaged in scan-and-vision meetings with prospective clients to understand their challenges and align them with tailored ML solutions.
- Contributed to public outreach on established conferences, for example on the topic of GANs and generative models at [PyconES23](#).

Visiting Deep Learning Engineer

2022 October – 2023 February

NVIDIA Corporation

- Carried out a research visit as part of a Marie-Curie Fellowship secondment.
- Worked on the design of AI-based super-resolution diffusion models for weather forecasting and the preparation of atmospheric variables (wind speed and direction at 100m). This effort was aligned with the Earth-2 digital twin project, relevant to the renewable wind energy sector.

Senior researcher in applied Data Science

2019 – 2023

Barcelona Supercomputing Center

- Led the applied AI research line within the Earth Sciences department.
- Managed technical teams (junior researchers, data scientists and engineers) to tackle different projects involving problems such as time series forecasting, image and video super-resolution, semantic segmentation, anomaly detection, pre-processing of big climate datasets (hundreds of terabytes), and gridded/raster data [visualization](#).
- Designed and implemented [DL4DS](#), a python open source package with AI-based super-resolution techniques (statistical downscaling) for climate/weather and EO gridded data.
- Presented my work in international conferences (e.g., GeoPython, [PyconES21](#), ESA Phi-week).
- Contributed to the writing of research proposals and supervised MSc students.
- Wrote peer-reviewed [research technical articles](#) and project deliverables.

Languages

- Spanish (native)
- English (professional)
- French (intermediate)
- Russian (intermediate)
- Catalan (beginner)

Awards and honors

- 2nd Prize at the [S2S AI challenge](#) organized by the WMO and ECMWF (2022)
- Marie Skłodowska-Curie COFUND fellowship at the Barcelona Supercomputing Center (2019-2023)
- ESA Copernicus accelerator (2020)
- 2nd Prize at the Barcelona [Copernicus hackathon](#) (2019)
- 1st Prize at the InvEnterPrize startup competition (2018)
- 1st Prize at the [ESA Phi-week startup bootcamp](#) (2018)
- Personal grant "Exoplanet direct imaging meets AI" (UGA, 2018)
- Scholarships for under- and post-graduate studies in Astrophysics by ICETEX/Colombia (2002) and CSIC/Spain International Campus of Excellence (2012)

Researcher in applied Data Science

2017 – 2019

Universite Grenoble Alpes (UGA), Grenoble Alpes Data Institute

- Implemented DL techniques for space sciences problems, such as the direct direction of extrasolar planets and the processing of satellite and Earth observation data.
- Supervised MSc and undergraduate students.
- Organized outreach and training activities (scientific programming and data science methods for scientists, [Software Carpentry @ UGA](#)).
- Presented my work in international conferences ([PyconES18](#)) and institutes (Caltech, Stanford).

Data science consultant

2017 March

Pivigo Ltd (Science to Data Science program)

- Developed a web scraping system, aimed at studying correlations between the text presentation, taken from UK schools websites, with the national ratings of school performance.
- Integrated a suite of NLP techniques, such as Tf-idf and sentiment analysis, alongside ML techniques including clustering, dimensionality reduction and ordinal regression. This system enabled in-depth investigation and delivery of key insights through a dashboard interface.

PhD researcher

2013 – 2017

Universite de Liege (STAR and Montefiore Institutes)

- Developed novel techniques for image processing and background subtraction: the [LLSG algorithm](#) (low-rank plus sparse decomposition of astronomical image sequences), and the [SODINN technique](#) for supervised detection of exoplanets using deep CNNs.
- Authored the [VIP python library](#) with a variety of ML-based image processing techniques under a clear API (docstrings, external documentation, Jupyter notebooks and CI with pytest/travis).

Junior researcher and lecturer

2009 – 2010

Technological University ITM, Colombia

- Taught the course of introductory Physics for engineers.
- Conducted research in astronomy and served as scientific advisor at the Planetarium "Jesús Emilio Ramírez González" of the city of Medellín, Colombia.

Skills

Soft

- Leadership and collaboration in multidisciplinary teams.
- Communication to non-technical stakeholders.
- Critical and scientific rigorous thinking.
- Willingness to study new topics and acquire new skills.
- Time organization to meet critical deadlines.

Hard

- Broad knowledge of statistics, ML, AI, and generative AI fields.
- Software development and software engineering.
- Ability to understand new data and to create preprocessing pipelines..
- Communication of complex technical concepts in a clear and accessible manner.
- Active learning attitude to stay up-to-date with new trends in the fields of ML and AI.

Technical

- [9+ years] Scientific Python ecosystem: Numpy, Scipy, Jupyter, Pandas, Xarray.
- [8+ years] Image processing and computer vision libraries: Opencv, Scikit-image, Pillow, Ultralytics.
- [8+ years] Machine learning and statistical analysis: Scikit-learn, Pycaret, Gluon, Autogluon, Xgboost, Catboost, Prophet, CausalML, Scikit-uptlift, Shap.
- [7+ years] Deep learning frameworks: Tensorflow/Keras, Pytorch, Cupy, HF-Transformers.
- [2+ years] LLMs and NLP: OpenAI-API, Langchain, Autogen, NLTK, Spacy, Tesseract OCR, Ollama.
- [4+ years] Project management: Jira, Trello, Confluence.
- [8+ years] Collaborative software development: Git, GitHub, GitLab.
- Development on clusters and cloud environments: Azure, HPC supercomputing clusters.
- Deployment and containerization: Docker.
- Quick dashboard/front-end development: Streamlit, Gradio
- Data visualization: Matplotlib, Bokeh, Plotly, Geoviews, Hvplot.
- Distributed ML workflows: Joblib, Horovod, Tf.distribute and Torch.distribute.
- Past experience in various languages: Fortran, C, .NET, Java, SQL, R, LaTeX.