Personal information & contact

- Nationality: Spanish, Colombian
- Location: Barcelona, Spain
- Telephone: +34 644092150
- Email: carlosgg33@gmail.com

- linkedin.com/in/carlosgog
- github.com/carlos-gg
- Google scholar
- 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

- Deployment of ML models on Azure cloud (2023, The Bridge)
- ESA bootcamp for EO 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)

Awards and honors

- 2nd Prize at the 2022 S2S AI challenge organized by the WMO and ECMWF
- Marie Sklodowska-Curie COFUND fellowship at the Barcelona Supercomputing Center
- ESA Copernicus accelerator 2020
- 2nd Prize at the 2019 Copernicus hackathon (Barcelona)
- 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)

Carlos Alberto Gomez Gonzalez

Lead data scientist

With over 7 years of extensive experience as a data scientist and AI engineer, I have cultivated a deep understanding of AI-driven innovation and the intricacies of spearheading 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-enabled projects in dynamic industries such as retail, manufacturing, and marketing within the private sector. My passion lies in harnessing cutting-edge technologies to deliver tangible solutions powered by the synergy of data science and AI.

Work Experience

Lead Data Scientist — 2023 June – Present

Decide4AI

- Independently developed internal R&D initiatives related to Causal ML and LLMs. Within the first initiative, a python library for uplift modeling was implemented. The line of work on LLMs resulted in a system for document/contract analysis integrating computer vision models (YOLO) for signature detection, OCR and text information retrieval using OpenAI models. The document analysis system featured a user-friendly Streamlit front-end and was efficiently containerized using Docker, enabling the auditing of internal contracts.
- Led the development of a document processing system for a Spanish startup, improving their legacy pipeline (built on Google cloud with AutoML models) with custom document classifiers and other NLP models to reduce the cost of inference by up to 50% and to improve metrics from 5% to 20% depending on the subtask and model.
- Led the creation of a credit scoring system for a prominent Chilean credit cooperative. Identified relevant metrics and enhanced performance up to 20% over their baseline model.
- Steered the development of an end-to-end solution tailored for a Chilean hygiene products manufacturer, aimed at forecasting sales volume up to 3 months in advance. This solution seamlessly integrated a demand elasticity model to enhance accuracy and strategic decision-making. The forecasting model had very good MAPE metrics from 5% to 40% (for the most difficult time series) and was containerized with Docker to facilitate its deployment.
- Directed the implementation of a demand forecasting system for a large international power tools manufacturer. This system accurately projected unit sales up to a 12-month horizon, leveraging a top-down reconciliation approach to enhance forecasts for distant horizons.
- Provided support to data science teams working with external clients (energy sector).
- 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 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.

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.

Languages

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

- 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).

Technician in Geographic Information Systems — 2011 – 2012

IGAC – Agustin Codazzi Geographical Institute, Colombia

- Participated in the program of digitalization of cadastral information using EO data and ArcGIS.

Junior researcher and lecturer — 2009 – 2010

Technological University ITM, Colombia

- Taught the course of introductory Physics for engineers.
- Carried out research in astronomy and acted as scientific advisor at the Planetarium “Jesús Emilio Ramírez González” of the city of Medellin, 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 and AI fields.
- Software development and engineering.
- Ability to clean and preprocess datasets.
- 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.
- [8+ years] Machine learning and statistical analysis: Scikit-learn, Pycaret, Gluon, Autogluon, Xgboost, Catboost, Prophet, CausalML, Scikit-uplift, Shap.
- [7+ years] Deep learning frameworks: Tensorflow/Keras, Pytorch, Cupy, Transformers (Huggingface).
- [2+ years] LLMs and NLP libraries: OpenAI-API, Langchain, NLTK, Spacy, Tesseract OCR.
- [4+ years] Project management: Jira, Trello.
- [8+ years] Collaborative software development: Git/GitHub/GitLab.
- Development on HPC clusters and cloud environments: Azure, AWS.
- Deployment and containerization: Docker.
- Quick dashboard/front-end development: Streamlit.
- Data visualization: Matplotlib, Bokeh, Plotly, Geoviews, Hvplot.
- Distributed ML workflows: Joblib, Horovod, Tf.distribute and Torch.distribute.
- Past experience in various languages: Fortran, C, C#, .NET, Java, SQL, R, LaTeX.