Miguel Cozar
Data Scientist

(401) 281-9695
Chicago, IL
miguel.cozar@icloud.com

Portfolio: mcozar99.github.io
github.com/mcozar99
linkedin.com/in/miguel-cozar

Highly skilled Data Scientist with experience in research, corporate data solutions and collaborative software development.

SKILLS

Tools and Languages	Python, Git, Excel, SQL, LaTeX, Java, JavaScript, MATLAB, React,
Machine Learning	Transformers, Torch, Keras, TensorFlow, Sklearn, Pandas
Cloud Platforms	Microsoft Azure, AWS, Google Cloud Platform, Terraform, Docker
Communication	English, Spanish (native), French (intermediate)

TECHNICAL EXPERIENCE

Data Scientist — October 2023 — Present
ServiceLink, a Fidelity National Financial company — *Remote, US*
- Fine tuned LLMs for real estate information extraction for 3000+ US counties.
- Led the first company's GenAI implementations for multiple use cases, with models like GPT-3.5, GPT-4, GPT-4o, PHI-3 or Llama-3.
- Designed optimization algorithms (Genetic Algorithm) for confidence tuning, increasing models' accuracy 10%+, while safely reducing false positives in production models.
- Devised labeling strategies with auto-agreement score scripts (via Cohen scores), helping the labeling team to boost their production of quality-labeled data.

Graduate Researcher / Computer Science Department — November 2021 — July 2022
Technical University of Madrid — *Madrid, Spain*
- Created and registered a webpage owner identification algorithm based on DNS routing and privacy policy scrapping (with Google Search Engine): helped to automatically identify 100+ companies behind web domains.
- Developed an algorithm utilizing Crunchbase, NetworkX, and Plotly to build comprehensive enterprise networks from a single company name, extracting and visualizing parent-child relationships to reveal complex connections between big corporations.
- Constructed an anycast, multicast and unicast IP geolocation algorithm based on multi-node network measures utilizing Python, Plotly and Mapbox.

Software Developer Intern — April 2019 — February 2020
Indra — *Madrid, Spain*
- Designed a radar emitter with python using Keras Neural Networks to simulate airplane radar emissions.
- Created a Softphone in bash using SIP protocol.

EDUCATION

Master of Artificial Intelligence, *Illinois Insitute of Technology* — 2023
- *Yield Engagement Mentor* (on-campus job): coached +10 new students to have a successful academic trajectory at Illinois Tech.
- *Connect Our Research Scholarship recipient*: led a coordinated project between IIT (Chicago) and UPM (Madrid) researching bias and fairness of machine learning models, resulted in a Master thesis with an A grade and an accepted paper in ICML 2023.

Master of Telecommunication Engineering, *Technical University of Madrid* — 2023
Bachelor of Telecommunication Engineering, *Technical University of Madrid* — 2021
- *Scholarship Recipient*: Completed Bachelor's final project (Design and Implementation of a Promotional Listening Service) with the Speech Technologies Group, experts in Natural Language Processing, resulting in an honors proposal.

ACTIVITIES

ICML 2023 Poster Presenter in the LatinXAI workshop	2023
Volunteer Content Creator for the LatinXAI Medium Page, inspiring other LatinX/Hispanic AI community members	2022
SECRYPT 2022 Poster presenter in main conference	2022
Euro S&P 2022 Research Speaker in IWPE Workshop, **Best Presentation of the day award**, voted by all attendants	2022

PROJECTS

- Developed an automatically deployed AWS hosted currency trading platform with Docker, Terraform and time-series databases (InfluxDB), obtaining maximum grade.
- Created an Open Source dataset with 150+ downloads of the US Senate network a deep analysis and description of the current US politic landscape.
- Devised a News Summarizer using BERT LLMs, deployed as a Twitter bot and able to answer user requests on stream
- Designed and developed Concierge, a web application for hotel management with React, JavaScript and Java using Scrum as a development methodology (5 developers)
- Built a song popularity predictor using Convolutional Neural Networks and Spotify API, demonstrating that lyrics are a key indicator of popularity.