Leopoldo Pla Sempere

✉ b64encoded:bGVvcG9sZG9wbGFzZW1wZXJlQGdtYWlsLmNvbQ== | ⌂ lpla.github.io | ▣ github.com/lpla

Skills

Scripting and tools: git, CMake, tmux, Python, Lisp, Bash, AWK, Perl, systemd, Snakemake, Docker
Machine Learning Tools: Keras, PyTorch, Hugging Face, pandas, sklearn, numpy, matplotlib
Programming: C++, C, Java, Golang, Rust, LaTeX, CMake, Lisp, PostgreSQL, Solr, PHP, CSS
Languages: English (Proficient), Spanish (Native), Catalan (Native)

Experience

Data Scientist, Treelogic, Remote   *Sep 2024 – Present*
- Research and development in projects as CRYPTOTRACK (monitor crypto transactions to detect illicit activities).

Part-time Lecturer, University of Alicante, San Vicente del Raspeig   *Oct 2019 – Present*
- Teaching in AI, Translation, Mathematics, and Computer Science degrees to over 200 students annually.
- Focus on programming fundamentals, data structures, and advanced software engineering techniques with real-world examples from leading tech companies and enhancing curriculum engagement through applied projects.

Innovation Projects Engineer, Grupo ENERCOOP, Crevillent   *Nov 2023 – Sep 2024*
- Lead engineer in Energy Production and Transition, managing national and EU-funded projects.
- Key AI and big data projects include DIGIRED-AI, InEExS, MODERATE, STREAM, and COMMUNITAS.

Senior NLP and Machine Learning Engineer, University of Alicante, San Vicente del Raspeig   *Nov 2017 – Nov 2023*
- Lead developer for the ParaCrawl and MaCoCu project, which processed over 1 PB of data, delivering public language models, parallel data, and web crawling and processing technologies that enhanced SOTA accuracies.
- Spearheaded the MultiScore project to develop music transcribing Transformer models.
- Managed teams of 7+ engineers, with international partners (JHU, UEDIN), handling HPC and cloud resources.

Data Scientist, NLP and Machine Learning Engineer, Prompsit Language Engineering, Elche   *May 2015 – Oct 2017*
- Key contributor to Reverso Context, handling more than 7 million of daily queries, enhancing search accuracy.

Projects

Bitextor, Bifixer, Bicleaner   github.com/bitextor
- Toolset to automatically harvest, align and clean texts from (multilingual) websites in any computing scale.

Publications and presentations

- Published research on web-scale acquisition of parallel corpora and marketable open data solutions.
- Speaker at **DockerCon 2022** and other technical showcases, discussing the Bitextor project impact.

Associations

- **Webmaster** for the European Association of Machine Translation (EAMT), technical services for eamt.org

Extracurricular activity

- Dedicated to **electronic repairs and coding for vintage video game platforms**, and actively supported community service efforts by 3D printing over 200 face masks during the COVID-19 pandemic.
- Professional musician, experienced in jazz, wind-orchestra, and rock performances.

Education

Polytechnic University of Valencia – M.Sc., Artificial Intelligence   *Sep 2014 – Jun 2015*
- GPA: 8.81/10. Specialized in Natural Language Processing and Deep Learning, with a thesis on Audio classical composer identification.

University of Alicante – B.Sc., Computer Science   *Sep 2010 – Jun 2014*
- GPA: 8.53/10. Focused on Informatics with a thesis on Dodecaphonic music composer assistant. Received the highest thesis defense mark (Outstanding with Honors) and awarded Best Academic Record.