Álvaro Peris

lvapeab@gmail.com • Website • Github • Google Scholar • LinkedIn

Machine learning engineer/data scientist. Problem solver. Demonstrated ability in tackling complex problems. Capable to own full, complex projects from their inception through their complete life cycle in a timely manner. As a result of working at startups, I'm able to wear many hats, and assume roles of software engineer, data scientist or data engineer.

I hold a PhD on machine learning, regarding deep learning for multimodal sequence-to-sequence modeling, human–computer interaction and adaptation.

WORK EXPERIENCE

CUSHION | Senior Machine Learning Engineer
2020 – Now

- **Email understanding**: Developed from scratch a system to extract insights from emails, featuring:
    - The first-ever data aggregator for BNPL purchases.
    - Detection of bills (utilities, insurance), subscriptions and credit card payments.
    - SKU-level data for retail purchases.
    - A reconciliation system to match data extracted from the email with transactions coming from other sources (bank statements, credit cards).
- This system processed 150K emails per day, with peaks of 300K/day. Emails were parsed by an in-house fine-tuned, quantized LLM. The overall automation rate was over 95%.
- Development and maintenance of advanced technology to tackle a variety of difficult problems related to customer finance, including:
    - **Recurring bill detection**: Built a state-of-the-art system to detect recurring bills from banks transactions. Developed a framework to discover and maintain missing bills through a combination of unsupervised learning and hand-crafted rules.
    - **Bank transaction categorization**: With a focus on building the best-in-class bank fee and refund detection and reconciliation.
    - **Overdraft prediction**: Implemented a system that warned our users of their risk of overdrafting their bank accounts.

SONIA LABS | Machine Learning Engineer
2019-2020

- Research and implementation of models for audio transcription and speaker diarization.
- Implementation of models to automatically parse and fill pdf forms.

PRHLT RESEARCH CENTER | Researcher
2015 – 2019

- Research on neural sequence-to-sequence models, with a focus on user-interaction, model adaptation (online and active learning) and multi-modality.
- Applied this research to machine transaction and image/video captioning.
- Authored +20 peer-reviewed papers, published in international conferences and journals.
- Open-sourced my research, providing toolkits to reproduce and extend the results.
- Supervised 5 MSc students who successfully defended their theses.

EDUCATION

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

PhD in Machine Learning
2015 — 2019

MSc in Artificial Intelligence, Pattern Recognition and Digital Imaging.
2014 — 2015

BSc in Computer Engineering
2008 — 2014

SKILLS

MACHINE LEARNING
Deep learning • LLMs • Classical ML
Active learning • Online learning
Sustainable ML cycles • Data annotation

SOFTWARE
Python • Django • SQL
PyTorch • Tensorflow • Transformers
Scikit-learn • Spacy • XGBoost

OTHER
Scientific communication • LaTeX
AWS • Grafana • Kibana • Docker

CODE PORTFOLIO

NMT-KERAS • DEMO
A fully-fledged neural machine transaction toolkit. The first and (back in the time) only open-source toolkit for neural machine transaction in Keras.

INTERACTIVE CAPTIONING • DEMO
Interactive multimedia captioning: Given an input image or video, interactively caption it.

FEB STATS • WEB (IN SPANISH)
A basketball stats parser and analyzer. Developed mostly to experiment with new tools; but used by professional Spanish basketball teams.