

CV // JOÃO GALEGO

[in linkedin.com/in/jgalego](https://www.linkedin.com/in/jgalego) | github.com/JGalego | jgalego.github.io

OVERVIEW

João Galego is a Principal AI Engineer at Critical Software and an Invited Professor at ISEG. He has a MSc in Physics from IST and a post-graduate degree in Forensic Studies from ISCSP. Before Critical Software, he worked as a DevOps Engineer and Lead ML Engineer at Siemens and as a Solutions Architect, ML Specialist, and Quantum Ambassador at AWS, helping startups at every stage of their cloud journey. When he is not working or teaching, you can usually find him browsing the local bookstore, contributing to open source projects, or enjoying some quality family time.

PROFESSIONAL EXPERIENCE

Principal AI Engineer

2025 – Present

Critical Software

- **Pre-Sales Support:** Collaborate with the sales team to understand client requirements, support the preparation of proposals, and present AI solutions. Communicate technical concepts to non-technical stakeholders, ensuring a clear understanding of AI projects and their impact.
- **Technical Leadership:** Oversee the end-to-end lifecycle of AI projects, from idea to production.
- **Solution Design:** Design and architect AI solutions that meet client needs and align with business goals.
- **Client Engagement:** Engage with clients to gather requirements, conduct demos, and provide technical expertise during the sales process.
- **Responsible AI Governance:** Ensure that all AI solutions comply with industry regulations, ethical standards, and company policies related to responsible AI development and deployment.
- **AI Research & Innovation:** Conduct cutting-edge research into emerging AI technologies and trends, evaluating their potential impact and integration into our AI solutions to provide a competitive edge.
- **Cross-Functional Collaboration:** Work with product management, engineering, data science, and other departments to deliver AI solutions that align with clients' objectives and business needs.
- **External Relationships:** Build and maintain partnerships with research organizations, academic institutions, and funding agencies to stay ahead of industry advancements and foster innovation opportunities.

Startup Solutions Architect

2022 – 2025

AWS

- Work directly with strategic Startup customers to accelerate challenging, mission-critical projects
- Create and share best practices, technical content and new reference architectures through white papers, code samples and blog posts
- Evangelize and educate about AWS technology through workshops, user groups, meetups, public speaking events, online videos and conferences
- Collaborate within the startup ecosystem (accelerators, incubators, VCs and meetups) to educate startups and support their growth
- Contribute to SA organizational growth by interviewing candidates and participating in hiring decisions
- Member of the EMEA South Startup SA team
- Full member of the AI/ML Technical Field Community (TFC) focused on MLOps
- Full member of the AI/ML Community of Practice (CoP)
- GenAI Subject Matter Expert (SME) for Startups
- GenAI Ambassador (L200)

- GenAI Expert (L300+)
- Inclusion Ambassador
- AI/ML Heroes program graduate
- AWS GameDay organizer, operator, maintainer and supporter
- AWS Senior Speaker Certified

Invited Professor

2024 – Present

ISEG / ULisboa

- Lecturer for the postgraduate program in Applied AI/ML
- Invited guest lecturer for the postgraduate program in AI for Value Creation

Research Collaborator

2020 – Present

EDGE Digital Lab

- Research initiative focused on the development and deployment of solutions for well-being, performance, longevity, and health based on exponential digital technologies such as AI, AR/VR, 3D Printing, Blockchain, 5G, Cloud, IoT, neuroprosthetics, and more.

Lead ML Engineer

2021 – 2022

Siemens

- Member of Siemens Financial Services' Artificial Intelligence and Automation (SFS IT&CYS AIA) chapter
- Development lead of robust ML solutions in an international and interdisciplinary DevOps team
- Senior subject-matter expert (SME) on ML technology at scale, deep learning (DL) and natural language processing (NLP)
- Responsible for deploying models into production while ensuring the highest standards in data integrity
- Responsible for implementing processes and systems for ML model lifecycle management

DevOps Engineer

2019 – 2021

Siemens

- Cloud solutions architect and administrator for 10+ applications running on AWS and Azure
- Lead contributor to 200+ Ansible roles, Terraform modules and Docker images
- Organized workshops and trainings on DevOps, AI/ML and quantum computing
- Performed technical assessment in 20+ interviews
- Supervised 10+ interns and trainees
- Junior member of the IT Key Experts program (Data-driven Enterprise & Augmented Intelligence)
- SITEX class of 2021 working on Industrial Edge platforms and applications
- Member of DI IT Architecture Guild

DevOps Engineer

2017 – 2019

InnoWave

- Designed and implemented test frameworks and automated tests for multiple projects and customers
- Organized internal and customer-oriented workshops and trainings on CI/CD, DevOps and test automation
- Served as technical writer and reviewer working together with software architecture teams

Test Automation Engineer

2014 – 2017

Coriant

- Member of the System Verification team for Optical Network Design & Planning tools
- Increased test automation coverage by 5-fold in less than 2 years
- Managed a Jenkins cluster with 20+ build agents

EDUCATION

PhD Cognitive Science (ABD)

2017 – 2021

ULisboa

Bringing minds and machines closer together using ML.

- **Title:** *From Thought to Action - Machine Learning Techniques for Processing Electrophysiological Signals in Brain-Computer Interfaces*
- **Advisors:** Prof. Hugo Ferreira (IBEB/FCUL), Prof. Arlindo Oliveira (INESC/IST)
- **Keywords:** motor cognition, brain-computer interfaces, biosignal processing, machine learning

PgD Forensic Studies 🎓

2016 – 2017

ISCSF

Learned what "every contact leaves a trace" actually means.

MSc Physics 🎓

2008 – 2013

IST

Failed to solve the riddle of baryon asymmetry.

- **Title:** *Flavour Aspects of Multi-Higgs Models* | **Grade:** 19/20
- **Advisors:** Prof. Gustavo Castelo Branco (CFTP/IST)
- **Keywords:** standard model, CP violation, discrete groups, flavour symmetries, multi-Higgs doublet models

CERTIFICATIONS



Oxford Machine Learning Summer School

2024

AI for Global Goals

ML summer school with dedicated tracks for ML x Health & Bio and ML x Rep. Learning & GenAI organised by AI for Global Goals in partnership with the University of Oxford's Deep Medicine and CIFAR and sponsored by DeepMind and GSK.

AWS Certified Solutions Architect - Professional

2024

Amazon

Professional-level exam intended for individuals with two or more years of hands-on experience designing and deploying cloud architecture on AWS.

AWS Certified Database - Specialty

2023

Amazon

Specialty-level exam to assess the candidate's comprehensive understanding of databases, including the concepts of design, migration, deployment, access, maintenance, automation, monitoring, security, and troubleshooting.

Oxford Machine Learning Summer School

2022

AI for Global Goals

ML summer school with dedicated tracks for ML x Finance and ML x Health organised by AI for Global Goals in partnership with the University of Oxford's Deep Medicine and CIFAR and sponsored by DeepMind and GSK.

Startup School	2022
Y Combinator	
7-week course featuring live talks by YC partners, founder stories by YC alumni, in-person meetups at YC offices all over the world, co-founder matching, and the largest community of startup founders on the internet cf. Announcing Startup School 2022 for the official announcement.	
AWS Certified Solutions Architect - Associate	2022
Amazon	
Associate-level exam to assess the candidate's ability to design and build secure, high-performing, cost-effective, highly available, and scalable solutions based on customer requirements.	
Migration Ambassador Foundations (Business)	2022
Amazon	
Earners of this badge are individuals who have demonstrated the foundational business knowledge about AWS Cloud Migration and the AWS Programs to support a customer's migration to the AWS Cloud.	
AWS Partner: Accreditation (Technical)	2022
Amazon	
This accreditation provides AWS Partners with fundamental, technical knowledge of AWS cloud computing, global infrastructure, services, common solutions, migration, security, and compliance.	
AWS Partner: Cloud Economics Accreditation	2022
Amazon	
Earners of this badge are AWS Partners who have demonstrated knowledge of cost savings and data center economics in relation to cloud computing.	
Machine Learning Engineering for Production (MLOps) Specialization	2021
DeepLearning.AI	
Multi-course specialization on how to <i>productionize</i> ML applications, organized in collaboration with DeepLearning.AI and the Google TensorFlow team.	
Natural Language Processing (NLP) Specialization	2021
DeepLearning.AI	
Multi-course specialization on how to design NLP applications that perform such tasks as sentiment analysis and language translation, organized in collaboration with DeepLearning.AI and the Google TensorFlow team.	
Google Data Analytics Professional Certificate	2021
Google	
Multi-course certificate program on how to collect, transform, and organize of data in order to draw conclusions, make predictions, and drive informed decision-making.	
AI Product Manager Nanodegree	2021
Udacity	
Course on how to develop AI products that deliver business value, organized in collaboration with Appen.	
Self-Driving Car Engineer Nanodegree	2021
Udacity	
Autonomous driving course covering several topics in computer vision, sensor fusion, localization, path planning, control and systems integration, organized in collaboration with Waymo and Mercedes-Benz.	
AWS Certified Machine Learning - Specialty	2020
Amazon	
Specialty-level exam to assess the candidate's ability to design, implement, deploy, and maintain ML solutions.	

Fundamentals of Digital Marketing	2020
Google	
Intro to digital marketing course accredited by Interactive Advertising Bureau Europe and The Open University.	
Microsoft Certified Azure Fundamentals (AZ-900)	2020
Microsoft	
Azure certification covering fundamental cloud concepts, Azure services, Azure workloads, security and privacy in Azure, as well as Azure pricing and support.	
Associate Android Developer	2018
Google	
Entry-level exam that includes a coding project and exit interview, covering topics related to app functionality, UI and data management, as well as debugging and testing of Android apps.	
Certified Jenkins Engineer	2016
Cloudbees	
Official certification covering core Jenkins concepts, Pipeline and Jenkins administration fundamentals.	
Machine Learning	2014
Coursera	
Introductory MOOC on machine learning taught by Andrew Ng.	

TECHNICAL SKILLS

Disclaimer: *This list is neither extensive nor exhaustive by design. As a firm believer in the “right tool for the right job” mindset, I’ve had the chance to work with hundreds of tools, packages and frameworks throughout my career. Listing them all here would be cumbersome and, worse, uninformative. Please contact me directly if you wish to discuss specific technical skills.*

Hands-on experience with multiple cloud providers (AWS, Azure, Google Cloud).

Active contributor to several open source projects:

- **AWS Samples** - Bedrock Access Gateway
- **AWS Labs** - Project Lakechain
- **Google DeepMind** - Concordia
- **Hugging Face** - Text Embedding Inference (TEI)
- **LangChain** - LangChain, LangServe
- **APR@NUS** - AutoCodeRover
- **Princeton NLP** SWE-Agent
- **ScrapeGraphAI** - Scrapegraph-ai
- **Stanford NLP** - DSPy
- **OpenGenerativeAI** - LLM Colosseum
- **DeCLaRe Lab** - LLM PuzzleTest
- **Aurelio AI Labs** - Semantic Router

Working knowledge of multiple programming and scripting languages:

- **Proficient** - Java, Python
- **Professional Use** - C++, C#, Dart (Flutter), Go, Groovy, JavaScript, Lua, MATLAB, Ruby, Rust
- **Academic Use** - C, Coq, Haskell, Lean, Lisp, Prolog, R, Wolfram Mathematica

Expert knowledge of multiple DevOps tools:

- **Automation** - Ansible, Packer, Terraform, Vagrant
- **CI/CD** - Azure DevOps, GitLab CI, Jenkins
- **Containerization** - Docker, Kubernetes
- **Logging & Monitoring** - Elastic (ELK) Stack, Grafana
- **Package Management** - JFrog Artifactory
- **Secrets Management** - HashiCorp Vault
- **Virtualization** - VirtualBox, VMware

Extensive experience with common Python packages for scientific computing (NumPy, SciPy), data analysis (Pandas, Vaex) and differentiable programming (JAX, MXNet, PyTorch, TensorFlow), NLP modules (Gensim, NLTK, spaCy, TextBlob), and high-level ML frameworks (HuggingFace Transformers and Diffusers, Keras, Scikit-Learn, Trax).

Expert knowledge of data and ML engineering tools and frameworks:

- **AutoML** - AutoGluon, AutoKeras NNI
- **Data & Model Versioning** - DVC, Git-LFS
- **Data Exploration** - JupyterLab, Jupyter Notebook
- **Data Processing** - Airflow, Hadoop, Spark 2.x and 3.x
- **Data Visualization** - Bokeh, Dash, Facets
- **Feature Storage** - Amazon SageMaker Feature Store, Feast
- **Hyperparameter Tuning** - Scikit-Optimize, Talos, Tune
- **ML Platforms** - Dataiku, Kubeflow, TFX
- **Model Interpretability** - Alibi, Lucid
- **Model Lifecycle** - MLflow
- **Model Serialisation** - ONNX
- **Model Serving** - TensorFlow Serving, TorchServe Triton Inference Server
- **Model Monitoring** - Evidently, Neptune
- **Optimization Tools** - Dask, Mahout
- **Simplification Tools** - Hermione, Ludwig
- **Visual Analysis and Debugging** - Manifold, Netron, Yellowbrick

HONORS & AWARDS

EMEA CS SA GenAI Builders Quest - 1st Place, GenAI Trailblazers

2023

AWS

Internal competition aimed to encourage SAs in EMEA CS to deepen their GenAI domain knowledge (L200+), supporting field readiness to guide customers with prescriptive guidance, advanced customer conversations, and earn their trust.

IBM Quantum Challenge - Advanced, 1st Place

2023

IBM

Quantum computing hackathon organized by IBM Quantum dedicated to dynamic circuits.

For more information, check out the official announcement on the IBM research blog.

- Hugging Face 🤗 Deep Reinforcement Learning Course - Certificate of Excellence** 2023
Hugging Face
Online course organized by Hugging Face covering beginner- to expert-level topics on Deep RL. Certificates of excellence were issued to all participants that passed **all** assignments before the end of April 2023.
For more information, check out the official announcement on the Hugging Face 🤗 blog.
- IBM Quantum Challenge - Advanced, 1st Place** 2022
IBM
Space-themed, quantum computing hackathon organized by IBM Quantum.
Exercises included an exploration of primitives using the new Qiskit Runtime, applying error mitigation techniques like matrix-free measurement mitigation (M3) to ML problems involving quantum kernels, implementing problem-specific parameterized quantum circuits to solve the traveling salesman problem, and the application of quantum chemistry techniques to study the interstellar medium by computing the reaction energy of H_3^+ , the first excited state and dipole moment of C_3H_2 (Cyclopropenylidene), and the interstellar chain reaction energy of $C + C_2H_2 \rightarrow C_3H_2$.
For more information, check out the official announcement on the IBM research blog.
- Machine Learning University - AutoGluon - Top 10** 2022
AWS
Internal AutoGluon competition promoted by Machine Learning University (MLU) as part of the *AutoGluon: Beyond the Basics* workshop - the goal was to create a model for binary classification on a multimodal setting (image, text, and tabular data).
- Machine Learning University - GML - Top 10** 2022
AWS
Internal graph machine learning (GML) competition promoted by Machine Learning University (MLU) as part of the accelerated GML track - the goal was to create a model to identify whether a chemical compound (represented as a graph) is *mutagenic* in nature or not cf. graph classification on MUTAG for additional information.
- QHack 2022 QML Challenge - Top 70 (*)** 2022
Xanadu
Quantum machine learning hackathon hosted by Xanadu and multiple sponsors.
Won 400\$ in credits for AWS and access to an IBM Quantum's 7-qubit device.
- (*) Due to conflicting engagements, participation in the Coding Challenge was only active during the 1st week.
- IBM QGSS21 Certificate of Quantum Excellence** 2021
IBM
Two-week intensive quantum computing summer school organized by the IBM Qiskit team with a focus on quantum machine learning. Quantum excellence certificates were issued to those who completed *all* graded lab work assignments with a final cumulative score above 75%.
For more information, check out the official announcement on the IBM research blog.
- Google Cloud Hero: BigQuery - Analytics & Machine Learning - 1st Place** 2021
Google / Siemens
Invite-only event organized by Google Cloud and hosted by the Google Siemens Account team. Google Cloud Hero is a *gamified* 3h training experience using hands-on labs focused on BigQuery and BigQuery ML.

IBM Quantum Challenge - Advanced, 1st Place

2021

IBM

Quantum computing hackathon organized by IBM to celebrate the 40-year anniversary of the *Physics of Computation* conference, and 5-year anniversary of IBM Quantum.

Exercises included factoring small integers with Shor's algorithm, running spectroscopy experiments against *real* IBM Quantum systems (w/ exclusive access to the brand new IBMQ Jakarta) using Qiskit Pulse and determining the ground state energy of small molecules (H_2 , LiH) using VQE.

For more information, check out the official announcement on the IBM research blog.

Udacity Technology Scholarship - AI Track

2021

Udacity / Bertelsmann

Full scholarship to attend Udacity's AI Product Manager nanodegree.

QHack 2021 QML Challenge - Top 50

2021

Xanadu / AWS

Quantum machine learning hackathon hosted by AWS and Xanadu.

Won 250\$ in credits for AWS and early access to an alpha version of Sandbox@Alphabet's Floq simulator 🦉 - a quantum SaaS API built on top of TPUs.

For more information, read my *QHack Recap - PennyLane, Amazon Braket and Beyond* article on Medium.

SITEX 2021 - Industrial Edge, 1st Place

2021

Siemens

Created an application that brings simulation (Tecnomatix® PlantSim) and optimization (Simcenter HEEDS) to the factory floor 🏭 and low code platforms (Mendix) to the Industrial Edge landscape as part of a strategic development program for the innovation and technology community in the Siemens hemisphere.

AI@Sustainability Hackathon - Technical Excellence (Team Mobility)

2020

Siemens

Built an application based on Amazon SageMaker in collaboration with *Swiss Federal Railways* (SBB) 🚆 to predict the *traffic volume* and *modal split* for a specific weather constellation and time based on a large set of real-world data points.

Associate Android Developer Fast Track Program Scholarship

2016

Google / Udacity

Full scholarship to attend Udacity's fast track course and take the Associate Android Developer exam.

Junior Research Scholarship

2013

Fundação para a Ciência e Tecnologia (FCT)

Scholarship for undergraduate students to work at the *Center for Theoretical Particle Physics* (CFTP).

ARTICLES & BLOGS

Hacking GraphRAG with Amazon Bedrock

2024

AWS Community

Learn how to run GraphRAG pipelines backed by Amazon Bedrock using LiteLLM proxy.

Make Programs, not Prompts: DSPy Pipelines with Llama 3 on Amazon SageMaker JumpStart

2024

AWS Community

Learn how to create ML pipelines with DSPy powered by Meta's Llama 3 70B Instruct model running on Amazon SageMaker.

Leaving no language behind with Amazon SageMaker Serverless Inference 2024

AWS Community

Host a high-quality translation model at scale using Amazon SageMaker Serverless Inference.

Build Document Processing Pipelines with Project Lakechain 2024

AWS Community

Learn how to create cloud-native, AI-powered document processing pipelines on AWS with Project Lakechain.

Your idea is my command: multi-agent collaboration for software development 2024

AWS Community

Create your own virtual software company with Amazon Bedrock and ChatDev and use it to build custom applications through LLM-powered multi-agent interactions.

Running LangChain.js Applications on AWS Lambda 2024

AWS Community

Learn how to run LangChain.js apps powered by Amazon Bedrock on AWS Lambda using function URLs and response streaming.

Scrape All Things: AI-powered web scraping with ScrapeGraphAI and Amazon Bedrock 2024

AWS Community

Learn how to extract information from documents and websites using natural language prompts.

Bug hunting with Amazon Bedrock and SWE-Agent 2024

AWS Community

Use Amazon Bedrock with SWE-Agent to create your own software engineering agent that can fix real-life bugs and issues in GitHub repositories.

Fighting Hallucinations with LLM Cascades 2024

AWS Community

Learn how to implement FrugalGPT-style LLM cascades on top of Amazon Bedrock using LangChain and LangGraph without breaking the bank.

Running Small Language Models on AWS Lambda 2024

AWS Community

Learn how to run small language models (SLMs) at scale on AWS Lambda with function URLs and response streaming.

Mapping embeddings: from meaning to vectors and back 2024

AWS Community

My adventures with RAGmap and RAGxplorer with a light introduction to embeddings, vector databases, dimensionality reduction and advanced retrieval mechanisms.

Building GenAI apps with managed AI services 2024

AWS Community

Some random thoughts on managed AI services and their place in the GenAI stack... with examples.

- The Rise of the LLM OS: From AIOS to MemGPT and beyond** 2024
 AWS Community
 A personal tale of experimentation involving LLMs and operating systems with some thoughts on how they might work together in a not-so-distant future.
- YOLOlambda: Running Serverless YOLOv8/9** 2024
 AWS Community
 Run inference at scale with YOLOv8/9 in a secure and reliable way with AWS Lambda.
- Deploy LangChain applications on AWS with LangServe** 2024
 AWS Community
 Learn how to deploy LangChain applications with LangServe in minutes on Amazon ECS and AWS Fargate using AWS Copilot.

TALKS & EVENTS

- O novo paradigma de GenAI com Amazon Bedrock** 2024
 AWS Generative AI Day
 L200 talk on Amazon Bedrock and GenAI for everyone.
- Demistifying ML with Amazon SageMaker: Agility, Speed and Efficiency** 2023
 AWS Cloud Experience Portugal
 L200 talk on low-code and no-code ML with Amazon SageMaker.
- How to solve common business challenges with AI/ML** 2022
 AWS Initiate Portugal 2022
 L200 talk featuring Sérgio Casaleiro (Software Development Manager at Ryanair Labs) on how to leverage the AWS ML Stack to solve common business problems and move an AI-powered application from PoC to production in no time.
- Amazon SageMaker for Startups** 2022
 AWS Experience Madrid on Floor 11
 L300 talk on how to use Amazon SageMaker to manage the entire MLOps pipeline at scale - from data exploration and feature engineering to model deployment and monitoring - with a focus on cost optimization.
- Machine Learning for Medicine and Healthcare** 2019
 12th Workshop on Biomedical Engineering (WBME)
 An overview of what ML actually is, highlighting some of its applications to the health sector - from the rise of expert systems in the 80s to the diagnosis, prognosis and treatment of COVID-19 - how it is shaping the present and how it may one day decide our future. Slides are available here

LANGUAGES

English (full professional proficiency), Español (professional working proficiency), Français (professional working proficiency), Italiano (elementary proficiency), Português (native)