João **Coelho**

MSc · Computer Science and Engineering · IST
PhD Candidate · Electrical Engineering/Language Technologies · IST/CMU
Lisbon, Portugal

(+351) 93 96 76 355 | joaomavc@gmail.com | jmvcoelho.github.io | JMVCoelho

Summary

João Coelho holds a MSc degree in Computer Science and Engineering from Instituto Superior Técnico (Lisbon, Portugal). He has experience working in research projects, and has published his work in international conferences and journals. He is interested in the areas of Machine Learning, Natural Language Processing, and Information Retrieval. He has experience programming in Python, including machine learning libraries such as PyTorch and Tensorflow, and is able to train and deploy Transformer models (BERT, GPT, T5, etc...) for multiple tasks. Other programming languages include C#, Java and C. He has also worked with both relational (MySQL, PostgreSQL) and non-relational (MongoDB) databases. João is currently a PhD candidate in a Dual-PhD program in Instituto Superior Técnico and Carnegie Mellon University. He is also a Teaching Assistant at Instituto Superior Técnico.

Work Experience

Caixa Mágica Software
Lisbon, Portugal

MACHINE LEARNING ENGINEER / RESEARCH
Sep. 2020 - Sep. 2022

- Project AppRecommender (PT2020 project number 39703)
  - **Description**: This project aimed to study novel approaches for mobile application search and recommendation.
  - **Tasks**: Scrapped a novel dataset containing metadata on mobile applications. Conducted data cleaning and analysis. Built a semantic search engine based on neural language models, leveraging a self-supervised training technique. Studied the impact of said model for *More Like This* recommendation. Created and deployed APIs to serve both the search engine and the recommendation system. Conducted two rounds of user-centered tests.
  - **Technologies**: Python, PyTorch, Flask, MongoDB, Docker.
- Other Responsibilities
  - Supervised four interns in two projects, (i) recommendation with collaborative filtering, (ii) slack chatbot.
  - Built and presented a large proof-of-concept to a client, regarding a project leveraging graph databases (Neo4J), domain-specific ontologies, semantic web, and neural models for entity extraction. The client accepted the proposal, resulting in one of the biggest projects held by the company.

Caixa Mágica Software
Lisbon, Portugal

INTERN DEVELOPER
Jul. 2020 - Sep. 2020

- Project Helpdesk
  - **Description**: This project aimed to build an helpdesk platform, where the questions that are supported by a FAQ section are automatically answered.
  - **Tasks**: Tested multiple approaches for FAQ retrieval under low resources. Integrated the best one within the previously built platform. Deployed the final product.
  - **Technologies**: Python, scikit-learn, Tensorflow, ReactJS, MySQL, Docker, Gitlab runners for CI/CD.

Xpand IT
Lisbon, Portugal

INTERN DEVELOPER
Jul. 2019 - Sep. 2019

- Project XBot
  - **Description**: This project aimed to build an internal proactive chat bot, which would send notifications or ask questions to the team.
  - **Tasks**: Implemented the backend of the bot, namely the proactive capabilities and possible message scenarios (notification, yes/no question, polls). Integrated a language understanding tool (LUIS). Deployed the final product in a Microsoft Teams channel.
  - **Technologies**: C#, .NET, Entity Framework, MySQL, Azure, LUIS.

Teaching

IST (Instituto Superio Técnico)
Lisbon, Portugal

TEACHING ASSISTANT
Sep. 2021 - Present

- Search and Planning (MEIC) - 2021/2022, 2022/2023
- Artificial Intelligence (LEIC) - 2021/2022, 2022/2023

Formal Education

CMU (Carnegie Mellon University)
Pittsburgh, PA, USA

PHD IN LANGUAGE TECHNOLOGIES
Sep. 2022 - Present

- Dual Degree PhD

IST (Instituto Superior Técnico)
Lisbon, Portugal

PHD IN ELECTRICAL AND COMPUTER ENGINEERING
Sep. 2022 - Present

- Dual Degree PhD

IST (Instituto Superior Técnico) *Lisbon, Portugal*

MSc in Computer Science and Engineering (Intelligent Systems/Language Technologies) [18.18/20] *Sep. 2019 - Nov. 2021*
- Merit Distinctions: 2019/2020 and 2020/2021.
- Thesis: *Geographical Question Answering Leveraging Neural Language Models for Passage Retrieval* [19/20]

IST (Instituto Superior Técnico) *Lisbon, Portugal*

BSc in Information Systems and Computer Engineering [15.46/20] *Sep. 2016 - Jun. 2019*
- Merit Distinctions: 2017/2018 and 2018/2019.

ESPS (Escola Secundária de Ponte de Sor) *Ponte de Sor, Portugal*

High School [18/20] *Sep. 2013 - Jun. 2016*
- Science and Technology Area.

Other Courses

European Summer School in Information Retrieval *Lisbon, Portugal*

Attendee *Jul. 2022*
- Week-long summer school focusing on Information Retrieval.

Deep Learning For Language Analysis Summer School *Koln, Germany (Virtual)*

Attendee *Aug. 2021*
- Week-long summer school focusing on Machine Learning for NLP.
- NVIDIA certificate: *Building Transformer-Based Natural Language Processing Applications*.

Lisbon Machine Learning Summer School *Lisbon, Portugal (Virtual)*

Attendee *Jul. 2021*
- Week-long summer school focusing on Machine Learning for NLP.

Publications

International Journals

1. João Coelho, Beatriz Paula, Diogo Mano, Carlos Coutinho, João Oliveira, Ricardo Ribeiro, and Fernando Batista, **Semantic Similarity for Mobile Application Recommendation under Scarce User Data**. Engineering Applications of Artificial Intelligence, 2023.

International Conferences

1. Beatriz Paula, João Coelho, Diogo Mano, Carlos Coutinho, João Oliveira, Ricardo Ribeiro, and Fernando Batista, **Collaborative filtering for mobile application recommendation with implicit feedback**. Proceedings of the IEEE ICE/ITMC Conference, 2022.

2. João Coelho, João Magalhães, and Bruno Martins, **Improving Neural Models for the Retrieval of Relevant Passages to Geographical Queries**. Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2021

3. João Coelho, António Neto, Miguel Tavares, Carlos Coutinho, João Oliveira, Ricardo Ribeiro and Fernando Batista, **Transformer-based Language Models for Semantic Search and Mobile Applications Retrieval**. Proceedings of the International Joint Conference on Knowledge Discovery and Information Retrieval, 2021.

4. João Coelho, António Neto, Miguel Tavares, Carlos Coutinho, Ricardo Ribeiro and Fernando Batista , **Semantic Search of Mobile Applications Using Word Embeddings**. Proceedings of the Symposium on Languages, Applications and Technologies, 2021.

Languages

English

C1 Level
- Certified by Oxford School
- Certified by Duolingo (140)

Portuguese

Native