

# Arian Pasquali

Machine Learning Engineer specialized in Generative AI evaluation with expertise in building evaluation methodologies for LLMs and AI agents. Experienced in bridging research with practical applications across government, education, and enterprise domains. Today my focus and main interest is building robust AI evaluation workflows for enterprise domains.

## EXPERIENCE

### Faktion — NLP Engineer

2022 - present. Antwerp, Belgium

Lead engineer specializing in generative AI evaluation and implementation across various sectors.

Core responsibilities:

- Lead Generative AI:
  - Architected LLM agents and multi-agent systems for education, construction, tax, legal, energy, supply chain, and government sectors
  - Implemented structured processes for systematic evaluation on AI Assistants
  - Built synthetic data generation solutions, data ingestion pipelines and AI automated report generation
- Evaluation-Driven Development applied to Generative AI:
  - Designed evaluation frameworks for generative AI applications for domain-specific use cases
  - Developed quantitative metrics that accurately predict user satisfaction, studying correlation between LLM-as-a-Judge with user preference in AB testing
  - Created methodologies to optimize domain expert time while maintaining data quality and confidence levels

### CitizenLab.co — NLP Engineer

2020 - 2022. Brussels, Belgium

Development of NLP API providing classification, clustering and information extraction services applied to multiple languages. Technologies applied: FastAPI as python web framework, Spacy and HuggingFace as NLP frameworks. Docker and Kubernetes as DevOps. AWS SageMaker as training and deployment infrastructure.

### INESC TEC Technology and Science — NLP Engineer

2020 - 2020. Porto, Portugal

Applied research engineering NLP solutions in the medical domain. Experience in tasks like Named Entity Recognition and Text Classification in Portuguese applied to medical records.

Technologies applied: Fast.ai and Spacy as NLP frameworks.

### Signal AI — Visiting Researcher

2019 - 2020. London, UK

Research and development in the field of text summarisation. My work has been focused on exploring and evaluating timeline summarisation methods from academia and applying them to news articles in a real world environment.

Antwerp, Belgium

(+32) 460 85 70 38

[arianpasquali@gmail.com](mailto:arianpasquali@gmail.com)

[github.com/arianpasquali](https://github.com/arianpasquali)

[www.linkedin.com/in/arianpasquali](https://www.linkedin.com/in/arianpasquali)

## SKILLS

Generative AI, Text Mining, Natural Language Processing, Machine Learning, Data Mining, Research

## ACADEMIC PUBLICATIONS

[Google Scholar Profile : Arian Pasquali](#)

## AWARDS

**Word Summit Award 2019 – Category Government & Citizen Engagement.**  
Founded in the framework of the United Nations' World Summit on the Information Society (WSIS), this award aims to promote the world's best innovative applications. My project [meuParlamento.pt](#) was awarded in the Citizen Engagement category.

**ECIR'19 Best Demo Presentation Award** (41st International Conference on Information Retrieval, Cologne – Germany) for the paper entitled "Interactive System for Automatically Generating Temporal Narratives" which presented a prototype for timeline summarization applied to web archives.

**ECIR'18 Best Short Paper Award** (40th International Conference on Information Retrieval, Grenoble – France) for the paper entitled "A Text Feature Based Automatic Keyword Extraction Method for Single Documents" which presented an unsupervised and language independent keyword extraction method (<https://github.com/LIAAD/yake>).

**Winner of the Portuguese Web Archive 2018 Awards** aims to award innovative research projects related to the Portuguese Web Archive (Arquivo.pt). Project Conta-me Histórias was awarded with the first prize.

## INESC TEC Technology and Science — Researcher

2016 – 2019. Porto, Portugal

Specialized in text mining at the Laboratory of Artificial Intelligence and Decision Support. Responsible for developing prototypes, proof of concepts, support on NLP initiatives and academic publishing. My main projects involved mining scientific publications, social media and news using python, Apache Solr or Elasticsearch.

## OSS PROJECTS

**YAKE** – <https://github.com/LIAAD/yake> (1.7k stars)

YAKE! is a widely adopted open source light-weight unsupervised automatic keyword extraction method which relies on text statistical features extracted from single documents to select the most important keywords of a text.

## EDUCATION

### University of Porto – Faculty of Sciences

*Master's degree in Computer Science with specialization in Data Mining*

2012 – 2015. Porto, Portugal

MSc thesis: Automatic Coherence Evaluation Applied to Topic Models [[pdf](#)]

## RELEVANT ACADEMIC PUBLICATIONS AND COLLABORATIONS

- 2020, YAKE! Keyword extraction from single documents using multiple local features, Information Sciences Journal.
- 2019, Interactive system for automatically generating temporal narratives, ECIR'19 (Best demo presentation paper).
- 2018, A text feature based automatic keyword extraction method for single, ECIR'18 (Best short paper).

## ADDITIONAL ACADEMIC COURSES

### Schloss Dagstuhl

*Autumn School of Information Retrieval and Information Foraging*

Summer School 2018. Wadern, Germany

Courses: Information Retrieval

### Universidad Politécnica de Madrid

*Advanced Statistics and Data Mining Summer School*

Summer School 2018 and 2019. Madrid, Spain

Courses: Text Mining, Big Data, Statistics and Deep Learning

### Universitat Pompeu Fabra

*European Summer School in Information Retrieval*

Summer School 2017. Barcelona, Spain

Courses: Information Retrieval

### Instituto Superior Técnico

*Lisbon Machine Learning Summer School*

Summer School 2012 and 2013. Lisbon, Portugal

Courses: Machine Learning focused on NLP

Winner of the Portuguese Web Archive

2019 Awards with the project, a mobile

application that explores the Portuguese

Parliament historical voting data.

## VOLUNTEERING

Reviewer in academic conferences and workshops:

– ECML PKDD 2025 – European

Conference on Machine Learning

– International Workshop on

Narrative Extraction from Texts

(2018 – 2025)

– ECIR 2020 – European Conference

on Information Retrieval

– International Workshop on User

Interfaces for Spatial and Temporal

Data Analysis (2019)

## CERTIFICATIONS

Google Cloud Generative AI Leader

AI Red Teaming Professional

Certification (AIRTP)

## LANGUAGES

English, Portuguese