Hugo Sousa

AI Researcher | Applied Scientist @ Amazon

hmosousa@gmail.com | hmosousa.com |

Research Interest

- Natural Language Processing
- Large Language Models
- Reinforcement Learning
- Information Retrieval

Highlighted Projects

Tradutor
Fine-tuned a language model to translate from English to European Portuguese.
code | model | app

tieval
A framework for evaluation of temporal information extraction systems.
code

Physio
A conversational AI agent to help with physical rehabilitation.
code

TEI2GO
Fast temporal expression identification systems.
code | models

Publications

Looking for the Bottleneck in Fine-grained Temporal Relation Classification
Hugo Sousa, Ricardo Campos, Alípio Jorge
SIGIR, 20-24 July 2026, Melbourne, Australia

The Temporal Game: A New Perspective on Temporal Relation Extraction
Hugo Sousa, Ricardo Campos, Alípio Jorge
CIKM, November 10-14, 2025, Seoul, Korea

Tradutor: Building a Variety Specific Translation Model

Hugo Sousa, Satya Almasian, Ricardo Campos, Alípio Jorge

AAAI, February 25 - March 4, 2025, Philadelphia, Pennsylvania, USA

Enhancing Portuguese Variety Identification with Cross-Domain Approaches

Hugo Sousa, Rúben Almeida, Purificação Silvano, Inês Cantante, Ricardo Campos, Alípio Jorge

AAAI, February 25 - March 4, 2025, Philadelphia, Pennsylvania, USA

Don't Forget This: Augmenting Results with Event-Aware Search

Hugo Sousa, Austin Ward, Omar Alonso

WSDM,10-14 March 2025, Hannover, Germany

Text2Story Lusa: A Dataset for Narrative Analysis in European Portuguese News Articles

Sergio Nunes, Alípio Jorge, Evelin Amorim, Hugo Sousa, António Leal, Purificação Silvano, Inês Cantante, and Ricardo Campos

LREC-COLING, 20-25 May 2024, Torino, Italy

Physio: An LLM-Powered Physiotherapy Advisor

Hugo Sousa, Rúben Almeida, Luís F. Cunha, Nuno Guimarães, Ricardo Campos and Alipio Jorge

ECIR, 24-28 March, 2024, Glasgow, Scotland

GPT Struct Me: Probing GPT Models on Narrative Entity Extraction

Hugo Sousa, Nuno Guimarães, Alípio Jorge, and Ricardo Campos

WI-IAT, 26-29 October 2023, Venice, Italy

TEI2GO: A Multilingual Approach for Fast Temporal Expression Identification

Hugo Sousa, Ricardo Campos, and Alípio Jorge

CIKM, 21-25 October 2023, Birmingham, United Kingdom

tieval: An Evaluation Framework for Temporal Information Extraction Systems

Hugo Sousa, Alípio Jorge and Ricardo Campos

SIGIR, 23-27 July 2023, Taipei, Taiwan

A Biomedical Entity Extraction Pipeline for Oncology Electronic Health Records in Portuguese

Hugo Sousa, Arian Pasquali, Alípio Jorge, Catarina Sousa, and Mário Lopes

ACM SAC, 27-31 March 2023, Tallinn, Estonia

Temporal Relation Extraction: The Event Ordering Task

Hugo Sousa

DESIRES, 15-18 September 2021, Padua, Italy

ECG Compression and QRS Detection: an IoT Approach

Hugo Sousa, Jonathan Tooley, Ana Paula Rocha

Master Thesis, 2019, Porto, Portugal

Talks

Principles for Developing Machine Learning Projects

Presented for master students in AI at the University of Porto

December 2025

Internships

Applied Scientist
Amazon
Jun 2024 — Sep 2024, Palo Alto

Visiting Student
Carnegie Mellon University
Oct 2023 — Dec 2023, Pittsburgh

Data Scientist
Microsoft
Feb 2019 — Apr 2019, Seattle

Education

Ph.D. Computer Science
University of Porto
Dec 2020 — Dec 2025, Porto

MSc. Applied Mathematics
University of Porto
Sep 2017 — Oct 2019, Porto

BSc. Physics
University of Porto
Sep 2014 — Jul 2017, Porto

Research Projects

CHPVVC
Sep 2023 — Present
The goal of this project is to produce a dashboard that compiles all the information regarding a patient to be used by the medical staff of the Hospital of Póvoa de Varzim and Vila do Conde (CHPVVC) that do Domiciliary visits. My role is to develop the natural language processing pipeline that will extract biomedical entities and relations between them from the multiple clinical documents.

StorySense
Jul 2023 — Present
The StorySense project is the sequel of Text2Story. While in the first project, we were concerned with extracting the information in a structured manner, in the StorySense project we are more focused on the meaning (sense) of each component of the narrative. My role in this project is as an assistant researcher, looking for a deeper encoding of temporal information.

MINE4HEALTH

Jan 2020 — Nov 2021

As a member of the NLP team, I contributed to the MINE4HEALTH project developed by INESC TEC in collaboration with the Portuguese Institute of Oncology of Porto - Francisco Gentil (IPO Porto). The project focused on using text-mining and natural language processing techniques to extract structured information from clinical documents. This effort aimed to enhance decision-making within oncological contexts by providing healthcare professionals with easily accessible panels containing crucial patient data, including health status, pathological history, biometric information, exams, and diagnoses. By streamlining the consultation process and enabling data-driven insights, the project not only reduces redundancy and treatment time but also supports the comparison of treatment effectiveness and adherence to international protocols.

Text2Story

Dec 2020 — Jun 2023

Member of the project Text2Story: Extracting journalistic narratives from text and representing them in a narrative modeling language. As a Ph.D. student, my contributions to this project were mainly concerned the extraction of temporal information from text. This includes the temporal entities - such as events, temporal expressions, or participants - and temporal relations that hold between them - such as before, after, or simultaneous.

Awards

3rd Sword AI Challenge

Member of the team - Team Physio - that took 3rd place on Sword AI Challenge 2023.

1st Portugal Digital Awards

Part of the team that developed the natural language processing pipeline for the MINE4HEALTH project. This project was recognized as the "Best Future of Work Project" at Portugal Digital Awards 2021.

2nd Eurekathon

Member of the team - Feeding the Future - that took 2nd place on Eurekathon 2020.

Employment History

Applied Scientist

Amazon

Jun 2025 — Present

Teaching Assistant

University of Porto

Feb 2023 — Present, Porto

Research Assistant

INESC TEC

Dec 2020 — Jun 2025, Porto

Data Scientist

BNP Paribas

Dec 2019 — Dec 2020, Porto

Data Scientist
JTA: The Data Scientist
Jul 2018 — Apr 2019, Porto

Other

Program Committee

- The 49th European Conference on Information Retrieval (ECIR'26)
- The 41st Annual AAAI Conference on Artificial Intelligence (AAAI'26)
- The 35th International Conference on Information and Knowledge Management (CIKM'26)
- The 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'25)
- The 48th European Conference on Information Retrieval (ECIR'25)
- The 40th Annual AAAI Conference on Artificial Intelligence (AAAI'25)
- The 34th International Conference on Information and Knowledge Management (CIKM'25)
- The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'25)

Reviewer

- 2 demonstration papers of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'24)
- 1 paper of the 47th International Conference on Theory and Practice of Digital Libraries (ECIR'24)
- 4 papers at the resource track of the 33nd ACM International Conference on Information and Knowledge Management (CIKM'24)
- 2 papers of the 27th International Conference on Theory and Practice of Digital Libraries (TPDL'23)
- 1 paper of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'23)
- 2 papers at the resource track of the 32nd ACM International Conference on Information and Knowledge Management (CIKM'23)
- 2 papers of the Sixth International Workshop on Narrative Extraction from Texts held in conjunction with the 45th European Conference on Information Retrieval (Text2Story'23 Workshop @ECIR)
- 1 paper of the 45th European Conference on Information Retrieval. ECIR'23
- 1 paper of the Web Conference (WWW'23)
- 1 paper of the IEEE Transactions on Knowledge and Data Engineering Journal in 2022
- 1 paper of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. (ASONAM'22)
- 1 paper of the 12th Temporal Web Analytics Workshop in conjunction with The Web Conference 2022 (TempWeb'22 Workshop @WWW)
- 1 paper of the Fifth International Workshop on Narrative Extraction from Texts held in conjunction with the 44th European Conference on Information Retrieval (Text2Story'22 Workshop @ECIR)
- 1 paper of the ACM Transactions on Information Systems Journal in 2021

- 1 paper of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'21)
- 2 papers of the 43rd edition of the annual BCS-IRSG European Conference on Information Retrieval (ECIR'21)

Chair

- Program Committee Member of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'25)
- Web Dissemination chair of the Eighth International Workshop on Narrative Extraction from Texts held in conjunction with the 47th European Conference on Information Retrieval (Text2Story'25 @ECIR)
- Session and Web Dissemination chair of the Seventh International Workshop on Narrative Extraction from Texts held in conjunction with the 46th European Conference on Information Retrieval (Text2Story'24 @ECIR)
- Session and Web and Dissemination chair of the 1st International Workshop on Implicit Author Characterization from Texts for Search and Retrieval held in conjunction with the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (IACT'23)
- Session and Web Dissemination chair of the Sixth International Workshop on Narrative Extraction from Texts held in conjunction with the 45th European Conference on Information Retrieval (Text2Story'23 @ECIR)
- Web Dissemination chair of the Fifth International Workshop on Narrative Extraction from Texts held in conjunction with the 44th European Conference on Information Retrieval (Text2Story'22 @ECIR)
- Session chair of the 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA'21)

Courses

- The 13th European Summer School in Information Retrieval, 18-22 July 2022, Lisbon (ESSIR'22)
- 11th Lisbon Machine Learning Summer School, 7-15 July 2021, Lisbon (LxMLS'21)
- Sequence Models with Attention Mechanism, July 2020, Online at Coursera
- Neural Networks and Deep Learning, May 2020, Online at Coursera