

# Andrei Stefan Bejgu (he/him)

**Date of Birth:** 16-11-1995

**Location:** Rome, Italy

**Languages:** English (Fluent), Italian (Native), Romanian (Native)

✉: stefan.bejgu@gmail.com

🌐: andreim14.github.io

in: andrei-stefan-bejgu

🐙 GitHub: github.com/andreim14

NLP professional with 7+ years of experience spanning software development, engineering, and research. I focus on multilingual NLP, retrieval-augmented generation (RAG), embedding-driven solutions and advanced Large Language Models (LLMs) fine-tuning – taking products from concept to large-scale deployment.

## EXPERIENCE & SELECTED PROJECTS

---

### Babelscape

*NLP Researcher*

*Rome, Italy*

Oct. 2021–Present

I lead advanced NLP research that transforms state-of-the-art insights into innovative AI products, with a focus on retrieval-augmented systems powered by multilingual LLM fine-tuning. I work closely with multidisciplinary teams to integrate these solutions, while mentoring interns at every stage of the project lifecycle.

- Designed and developed real-time retrieval pipelines, maximizing precision and efficiency for Vera, a multilingual fact-checking system.
- **Distributed** (Slurm) fine-tuning of open-source LLMs for domain-specific tasks such as **claim extraction and verification**.
- Contributed to the safety fine-tuning of **Minerva**, the first Italian LLM.
- Engineered **Multilingual Embeddings** for RAG systems, driving state-of-the-art retrieval.
- Improved **Entity Linking** in Comprehendo, achieving a **12%** accuracy gain while reducing overhead by **60%**.

*NLP Engineer*

Jun. 2020–Oct. 2021

I built robust, production-grade solutions by developing scalable pipelines that integrate technologies like Docker, Kubernetes, and Spark. I also implemented efficient data processing and indexing systems to ensure high performance and reliability across various AI products.

- Designed and deployed Semantic Paths, a **Multilingual Semantic Search** engine for scientific articles written in Kotlin.
- Developed advanced **Word Sense Disambiguation** models for large-scale multilingual processing.
- Optimized NLP systems for speed, accuracy, and scalability, improving overall robustness and reliability.

*Software Engineer*

Jul. 2017–Jun. 2020

I engineered scalable, high-performance systems by designing and implementing multiprocess pipelines for enterprise-grade applications.

- Developed myKnowledgeGraph, a **Knowledge Graph Extraction** and visualization system for enterprise data.
- Designed the company's **Java NLP Pipeline**, maximizing speed, accuracy, and scalability of multilingual text processing.
- Enhanced **concept identification** in TraDeInterpret, increasing accuracy by **13%** in trademark analysis.

### CINI - Consorzio Interuniversitario Nazionale per l'Informatica

*Software Developer*

*Rome, Italy*

Dec. 2016–Jan. 2017

Created an advanced tool for lexicon-semantic stylometric analysis, enabling efficient processing of large-scale texts.

## EDUCATION

---

### Ph.D. in Artificial Intelligence

*Sapienza University of Rome*

Jan. 2025

**Focus:** Explored multilingual NLP by applying lexical semantics to modern systems. Improved domain-specific tasks, looked into complex linguistic patterns, and worked on word- and sentence-level representations in RAG systems and LLMs. This was a joint industrial Ph.D., combining academic research with industry-driven projects. Created context-aware, domain-specific embeddings that boosted retrieval performance. –**Advisor:** Prof. R. Navigli

### MSc in Computer Science

*Sapienza University of Rome* –**Grade:** 110/110 with honors

Nov. 2021

**Thesis:** Exploring language-agnostic latent spaces for sentences. –**Advisor:** Prof. R. Navigli

### BSc in Computer Science

*Sapienza University of Rome* –**Grade:** 107/110

Jun. 2019

**Thesis:** Knowledge-Based Word Sense Disambiguation leveraging synset embeddings. –**Advisor:** Prof. R. Navigli

## PUBLICATIONS

---

- [1] A. Scirè\*, **A. S. Bejgu\***, S. Tedeschi, K. Ghonim, F. Martelli, and R. Navigli. “Truth or Mirage? Towards End-to-End Factuality Evaluation with LLM-OASIS”. In: *Under Review*. 2025.
- [2] L. Moroni, G. Puccetti, P.-L. H. Cabot, **A. S. Bejgu**, A. Miaschi, E. Barba, F. dell’Orletta, A. Esuli, and R. Navigli. “Adaptation of LLMs toward Italian Language: A Systematic Approach”. In: *NAACL*. 2025.
- [3] K. Ghonim, **A. S. Bejgu**, and R. Navigli. “Conceptpedia: Leveraging Traditional Approaches for Multimodal Image-Text Representation”. In: *Under Review*. 2025.
- [4] **A. S. Bejgu**, E. Barba, L. Procopio, A. Fernández-Castro, and R. Navigli. “Word Sense Linking: Disambiguating Outside the Sandbox”. In: *ACL*. 2024.
- [5] F. Molfese, **A. S. Bejgu**, S. Tedeschi, S. Conia, and R. Navigli. “CroCoAlign: A Cross-Lingual, Context-Aware and Fully-Neural Sentence Alignment System for Long Texts”. In: *EACL*. 2024.
- [6] F. Martelli, **A. S. Bejgu**, et al. “XL-WA: a Gold Evaluation Benchmark for Word Alignment in 14 Language Pairs”. In: *CLiC-it*. 2023.

## OTHER ACTIVITIES

---

**Reviewer Activity:** Reviewer for EACL (2023), ACL (2023, external reviewer), EMNLP (2023, external reviewer), ICLR (2024, external reviewer), and ACL ARR (January, June, and October 2024).

**Teaching Activity at Sapienza University of Rome:**

- **Natural Language Processing Course (2022–2023):** Designed and led practical lab sessions and assignments, guiding master’s students through state-of-the-art NLP techniques.
- **Hot Topics in NLP (2023–2024):** Developed and managed one of the three major course assignments, providing hands-on exposure to cutting-edge advancements in NLP.

## SCHOLARSHIPS AND AWARDS

---

- Best Employee Award, Babelscape 2021
- Merit Scholarship Winner, LazioDiSCo.it 2018–2020