João Augusto Leite

Ph.D. Candidate, Research Associate
The University of Sheffield

✉ j.leite@sheffield.ac.uk
⌨ JAugusto97
🔗 joao-augusto-leite
🌐 jaleite.com

EXPERIENCE

The University of Sheffield   *JUNE 2024 - Present*
(Part-time) Research Associate   Sheffield (UK)
- Developing content verification tools for the Chinese language.

The University of Sheffield   *SEPT 2023 - JAN 2024*
(Part-time) Graduate Teaching Assistant - Text Processing Module (3110/4115)   Sheffield (UK)
- Prepared and presented lab demonstrations to undergraduate students.
- Marked undergraduate work.

PicPay   *OCT 2021 - OCT 2022*
Data Scientist   São Paulo (BR) - Remote
- Designed and served information retrieval models for search engines used by millions of users.
- Conducted A/B test experiments for multiple POCs.

Birdie.ai   *AUG 2019 - OCT 2021*
Machine Learning Engineer   São Carlos (BR)
- Designed and implemented *SOTA* NLP models for the data enrichment pipeline on a variety of applications: NER, Sentiment Analysis, Clustering, Disambiguation, Ontology Building.
- Built web crawlers to collect product reviews from major retailers such as Amazon and Ebay.
- Designed and led labelling tasks with a dedicated team to serve supervised machine learning models.

EDUCATION

University of Sheffield (United Kingdom)   *Nov 2022-May 2026 (Expected)*
Doctor of Philosophy, Computer Science

Universidade Federal de São Carlos (Brazil)   *2021-2024*
Master of Science, Computer Science

Universidade Federal de São Carlos (Brazil)   *2017-2021*
Bachelor of Science, Computer Science

PUBLICATIONS

Asterisk (*) indicates equal contribution.

Social Media Content Credibility Assessment: A Survey on Automatic Detection [...]   *2024*
Pre-print | I. Srba, O. Razuvayevskaya, **J.A. Leite**, R. Moro, I.B. Schlicht, S. Tonelli, F.M. Garcia, S.B. Lottmann, [...]

EUvsDisinfo: a Dataset for Multilingual Detection of Pro-Kremlin Disinformation in News Articles   *2024*
CIKM | **J.A. Leite**, O. Razuvayevskaya, K. Bontcheva, C. Scarton

Detecting Misinformation with LLM-Predicted Credibility Signals and Weak Supervision   *2023*
Pre-print | **J.A. Leite**, O. Razuvayevskaya, K. Bontcheva, C. Scarton

Comparison between parameter-efficient techniques and full fine-tuning   *2023*
Plos One | O. Razuvayevskaya*, B. Wu*, **J.A. Leite***, F. Heppell*, I. Srba, C. Scarton, K. Bontcheva, X. Song

Mono and multilingual approaches for news genre, topic and persuasion technique classification   *2023*
ACL-SemEval | B. Wu*, O. Razuvayevskaya*, F. Heppell*, **J.A. Leite***, C. Scarton, K. Bontcheva, X. Song

Noisy Self-Training with Data Augmentations for Offensive and Hate Speech Detection Tasks   *2023*
RANLP | **J.A. Leite**, C. Scarton, D.F. Silva

Toxic Language Detection in Social Media for Brazilian Portuguese   *2020*
AACL-IJCNLP | **J.A. Leite**, D.F. Silva, K. Bontcheva, C. Scarton