Atilla Kaan ALKAN

Ph.D. | Astrophysical NLP, Statistical Learning, Information Extraction

atillakaan.alkan@gmail.com +33 6 51 47 37 37 LinkedIn
aalkan Google Scholar orcid.org/0000-0001-7964-4420

Experience

Research

Since September 2024 — **Research Member,** UniverseTBD.
Research in Natural Language Processing (NLP) for Astrophysics.

July 2023 – September 2023 — **Visiting PhD Candidate,** Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.
Designed and implemented an information extraction pipeline for astronomy publications at NASA Astrophysics Data System, using a wide range of NLP techniques, including named entity recognition, coreference resolution, entity linking/normalization, and relation extraction. Work is done in Python and is openly available on GitHub.

November 2020 - April 2021 — **Research Intern,** CEA-Saclay, Institute of Research into the fundamental Laws of the Universe (Irfu).
Implementation from scratch of Word2Vec (with NumPy) for astrophysical term representations.

March 2020 - August 2020 — **Research Intern,** French Ministry of Home Affairs, Paris.
NLP techniques for improving *HistoVec*, a French Home Affairs Ministry tool: Implementation of a semantic similarity-based clustering with pre-trained language model CamemBERT and the Latent Dirichlet Algorithm for topic modelling.

June 2019 - August 2019 — **Research Intern,** French Ministry of Home Affairs, Paris.
Implementation of statistical learning methods (decision trees and random forests) for predicting the number of fire brigade interventions: development of a daily prediction model.

Teaching

September 2024 – January 2025 — **Statistical Learning Lecturer,** Polytechnic Institute to Advanced Sciences, Paris.

- Mathematical Foundations for Statistical Learning (32 hours)
- Hands-On Machine Learning for Cybersecurity (20 hours)
- Mathematical Analysis (Lab sessions only - 20 hours)

Experience (continued)

February 2023 – June 2024 — **Computer Science Teaching Assistant,** Technological University of Orsay (IUT d'Orsay).

- Object Oriented Programming in Java (32 hours)
- Procedural Language extensions to the Structured Query Language (PL/SQL) (58 hours)

Education

2021 – 2024 — **Ph.D. in Astrophysical Natural Language Processing,** Paris-Saclay University.
Thesis title: *Natural Language Processing for Analyzing Messages of Astrophysical Observations.*

2020 – 2021 — **Specialization in Statistical Learning,** Polytechnic Institute of Advanced Sciences, Paris.
Architecture of recurrent and convolution neural networks, optimisation strategies, neural language models.

2017 – 2020 — **M.Sc. Aeronautics and Aerospace,** Polytechnic Institute of Advanced Sciences, Paris.
Studied fundamental astronomy, astrophysics, data analysis methods and modelling.

Sep. 2018– Dec. 2018 — **International Semester,** University of Limerick, Ireland.
Aircraft Flight Dynamics and Simulation, Aerospace Structures.

Publications

Conference Proceedings

1. **Alkan, Atilla Kaan**, F. Grezes, C. Grouin, F. Schüssler, and P. Zweigenbaum, "Astroecr : Enrichissement d'un corpus astrophysique en entités nommées, coréférences et relations sémantiques.," in *Actes de la 31e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, Toulouse, France, 2024.

2. **Alkan, Atilla Kaan**, F. Grezes, C. Grouin, F. Schüssler, and P. Zweigenbaum, "Enriching a time-domain astrophysics corpus with named entity, coreference and astrophysical relationship annotations," in *Proceedings of the Fourteenth Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*, Torino, Italy: European Language Resources Association, May 2024.

3. **Alkan, Atilla Kaan**, C. Grouin, and P. Zweigenbaum, "Étude de méthodes d'augmentation de données pour la reconnaissance d'entités nommées en astrophysique," French, in *Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 1 : travaux de recherche originaux – articles longs*, Paris, France: ATALA, Jun. 2023, pp. 1–13. URL: https://aclanthology.org/2023.jeptalnrecital-long.1.

4. **Alkan, Atilla Kaan**, C. Grouin, F. Schussler, and P. Zweigenbaum, "A majority voting strategy of a SciBERT-based ensemble models for detecting entities in the astrophysics literature (shared task)," in *Proceedings of the first Workshop on Information Extraction from Scientific Publications*, Online: Association for Computational Linguistics, Nov. 2022, pp. 145–150. URL: https://aclanthology.org/2022.wiesp-1.17.

5. **Alkan, Atilla Kaan**, C. Grouin, F. Schussler, and P. Zweigenbaum, "TDAC, the first corpus in time-domain astrophysics: Analysis and first experiments on named entity recognition," in *Proceedings of the first Workshop on Information Extraction from Scientific Publications*, Online: Association for Computational Linguistics, Nov. 2022, pp. 131–139. URL: https://aclanthology.org/2022.wiesp-1.15.

6. **Alkan, Atilla Kaan**, V. Lefranc, P. Reichherzer, and F. Schüssler, "Transient astronomical phenomena monitoring with astro-colibri.," in *SF2A-2022: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, Online, 2022. URL: https://ui.adsabs.harvard.edu/abs/2022sf2a.conf..135A.

7. F. Belbachir and **Alkan, Atilla Kaan**, "Features influencing the concept of trust in online reviews," in *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, 2022, pp. 1–6. DOI: 10.23919/CISTI54924.2022.9820186.

Journal Articles

1. P. Reichherzer, F. Schüssler, V. Lefranc, J. Becker Tjus, J. Mourier, and **Alkan, Atilla Kaan**, "Astro-colibri 2; an advanced platform for real-time multi-messenger discoveries," *Galaxies*, vol. 11, no. 1, 2023, ISSN: 2075-4434. DOI: 10.3390/galaxies11010022.

2. P. Reichherzer, F. Schüssler, V. Lefranc, *et al.*, "Astro-colibri—the coincidence library for real-time inquiry for multimessenger astrophysics," *The Astrophysical Journal Supplement Series*, vol. 256, no. 1, p. 5, Aug. 2021. DOI: 10.3847/1538-4365/ac1517.

Popularization Book Co-Author

1. J.-M. Jaspers, **Alkan, Atilla Kaan**, V. Beloulou, T.-R. Tabi-Agbor, and M. Simon-Sani, *Le traitement automatique du langage, le "tal" en 40 chapitres.* 2022.

Additional Information

Languages

- French (native), Turkish (native), English (fluent), Spanish (conversational), Arabic (basics).

Certification

2018 **Aeronautical Teaching Certificate,** Academy of Creteil.
Teaching basics of aeronautics aeroclub students and exam preparation.

2015 **Aeronautical Initiation Certificate,** Aeroclub of Alsace, Strasbourg.
Fundamental of aeronautics: flight dynamics, meteorology, aviation regulation and navigation.

Awards

2021 **AISIS 2021 prize for an outstanding contribution of relevance to Science.** The Second edition of Artificial Intelligence for Science, Industry and Society (AISIS), National Autonomous University of Mexico (UNAM).

Science Outreach and Community Service

2024 **Acadomia Private Tutoring,** Group tutoring in Mathematics for final year high school students preparing for the French Baccalauréat exam.

Additional Information (continued)

2023 **Organization committee member,** The Second Edition of the UDOPIA PhD Students Day, CentraleSupélec, December 2023.

Popularization Book Co-Author. The book aims to introduce the field of Natural Language Processing (NLP) to the civil servants of the French Home Affairs Ministry.

2022 **Organization committee member,** LISN Laboratory PhD Students Day, LISN, June 2022.

2021 **Organization committee member,** The First Edition of the UDOPIA PhD Students Day, CentraleSupélec, December 2022.