

# Tristan DONZÉ

Master 2 Student | Seeking end-of-studies internship from April 2026 (6 months)

 [www.tristandonze.fr](http://www.tristandonze.fr)  [tristan.donze@polytechnique.edu](mailto:tristan.donze@polytechnique.edu)  [linkedin/tristandonze](https://linkedin/tristandonze)  [github/TristanDonze](https://github/TristanDonze)

## EDUCATION

### Institut Polytechnique de Paris

Research Master's in AI (English-taught) | Courses delivered at *École Polytechnique & Télécom Paris.*

Sep. 2024 – Sep. 2026

Palaiseau, France

### Université Paris 8

Computer Science Bachelor's degree. Graduated with high honours.

Sep. 2021 – Jun. 2024

Saint-Denis, France

## TOPICS COVERED

- Deep Learning
- Deep Reinforcement Learning

- Computer Vision
- Natural Language Processing

- Graph Machine Learning
- Generative Modeling

## EXPERIENCE

### Histia

Station F, Paris, France

Freelance AI Engineer | *Agent Orchestration, Google Agent Development Kit*

Sep. 2025 – Present

- Engineered an autonomous multi-agent pipeline that crawls corporate websites, analyzes both markdown-converted text and page-level visual data, and generates high-fidelity company reports suitable for VC, incubator, and M&A workflows.
- Built parallelized page-analysis components that independently process and summarize textual and visual signals, along with a fusion module that merges their outputs into unified, page-level insights.
- Developed an adaptive orchestration mechanism that selects the next pages to explore based on missing sections in the evolving report, improving coverage and reducing redundant traversal.
- Implemented system-level optimizations reducing token usage by ~40%, achieving strong throughput and reliably capturing the full scope of a company's activities.

### Applied Research Intern | *Vision-Language Models, PyTorch*

Apr. 2025 – Aug. 2025

- Designed and implemented an end-to-end system for company logo detection and identification in both natural images and isolated logo samples.
- Curated and standardized a large-scale dataset of ~3M logo instances, and built a manually annotated test set aligned with real-world evaluation needs.
- Fine-tuned the CLIP image encoder with contrastive learning and LoRA adapters, improving model robustness through iterative error analysis and targeted data augmentations.
- Achieved 94.6% Top-1 accuracy on the test set, surpassing the baseline by more than 10 percentage points.

## PROJECTS

### Deep Reinforcement Learning Blackbox Challenge | *PPO, Noisy Environment, PyTorch*



- Engineered agents to tackle a fully unknown, highly noisy environment with no access to underlying dynamics or documentation, operating under strict neural network capacity constraints (128 neurons maximum).
- Implemented A2C and PPO algorithms with Generalized Advantage Estimation (GAE), adaptive entropy scheduling, KL-based early stopping, and cosine annealing. Designed a shared encoder architecture to maximize representational efficiency.
- Ranked among the top performers in the course challenge, achieving an average reward of 50 across 100 evaluation episodes.

### Political Speech Imitation & Fallacy Detection | *LLM Fine-tuning, QLoRA, Mistral, Phi4, Rhetoric Analysis*



- Fine-tuned multiple LLMs (Phi4-mini, Mistral-7B/24B) using QLoRA on two tasks: (1) imitating rhetorical styles from political speech corpora (630 speeches transformed into 16k instruction-response pairs via Gemini API), and (2) detecting logical fallacies from 5.7k labeled arguments across 9 fallacy types.
- Observed that rhetoric-adapted models produced significantly more fallacious reasoning compared to baseline LLMs; evaluated using BERTScore, binary classification, and linguistic feature analysis.
- Benchmarked 6+ models on automated fallacy detection across 9 fallacy types (Ad Hominem, Strawman, Appeal to Authority, etc.), analyzing precision-recall tradeoffs and per-class performance variations.

### Football Event Detection from Tweets | *NLP, Time Series Classification, BERT, TCN, XGBoost*



- Developed a system to identify key football events (goals, cards, etc.) during 2014 World Cup matches by classifying multilingual Twitter streams in real time. Explored multiple approaches: BERT fine-tuning (LoRA, P-Tuning), embedding-based classifiers (XGBoost, logistic regression), MLPs using word importance matrices, and Temporal Convolutional Networks.
- Achieved highest accuracy with MLP (78%) and TCN (75%) models for event detection.

## TECHNICAL SKILLS

**Programming Languages :** Python, C, C++, SQL

**Libraries & Frameworks :** PyTorch, HuggingFace, NumPy, Agent Development Kit (ADK), Crawl4AI

**Development Tools :** Linux, GitHub, SLURM, Aim

**Languages :** English (Fluent), French (Native), Italian (Learning)

## ACCOMPLISHMENTS

**Hackathon :** Second place at the Tech: Europe hackathon (90 participants, 30 teams) with **MagnOSS**, an open-source chess coach leveraging AI for personalized game analysis, interactive theory exploration, and Chess.com integration. 2025

**Open-Source :** Contributed to Crawl4AI by proposing and implementing a new feature to enhance its functionality. 2025

**Volunteer Program :** Participated in a multicultural volunteer program in Iceland, where I contributed to the design and construction of urban furniture in a natural setting. Collaborated with an international team to create a pedestrian-friendly area. 2023

## EXTRACURRICULAR ACTIVITIES

**Music :** Learning guitar, practicing mixing, and regularly playing the piano.

**Chess :** Passionate player, training regularly to improve.

**Hiking :** Enthusiast, planning to climb Mont Blanc soon.