Tristan DONZÉ

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

www.tristandonze.fr tristan.donze@polytechnique.edu linkedin/tristandonze github/TristanDonze

Education

Institut Polytechnique de Paris *Sep. 2024 – Sep. 2026*
Research Master's in AI (English-taught) | Courses delivered at *École Polytechnique & Télécom Paris* *Palaiseau, France*

Université Paris 8 *Sep. 2021 – Jun. 2024*
Computer Science Bachelor's degree. Graduated with high honours. *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*

- Developed an autonomous multi-agent system that explores company websites, analyzes both textual content (via markdown conversion) and visual context (via screenshots), and generates structured, investment-grade reports for use by VC funds, incubators and M&A firms.
- Designed a parallel processing pipeline where specialized agents independently summarize textual and visual information per page, and a synthesis component merges their outputs into cohesive insights.
- Implemented a coordination mechanism that dynamically selects which pages to analyze next based on missing information in the report, improving coverage and report completeness.

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*

- Built agents to solve a fully unknown environment with highly noisy observations and no access to dynamics or documentation, under strict constraints on neural network capacity (128 neurons max)
- Implemented A2C and PPO with GAE, entropy scheduling, KL early stopping, and cosine annealing. Designed a shared encoder architecture optimized to maximize representational efficiency
- Achieved top tier performance in the course challenge, demonstrating strong sample efficiency and robustness

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*

- Detected key moments (goals, cards) during 2014 World Cup matches from multilingual Twitter streams using binary time-period classification. Experimented with multiple approaches: BERT fine-tuning (LoRA, P-Tuning), embedding-based classifiers (XGBoost, logistic regression), and Temporal Convolutional Networks
- Achieved best performance with TCN architecture by capturing temporal patterns in tweet embeddings aggregated over match periods, extracting reliable signals from massive, noisy, multilingual text data

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.