Baptiste PRAS

Master's student in Artificial Intelligence

✉ baptiste.pras@universite-paris-saclay.fr 🔗 baptistepras.fr/ ⌗ https://github.com/baptistepras

in https://www.linkedin.com/in/baptiste-pras/

EDUCATION

Master's degree in Artificial Intelligence
University Paris-Saclay

Sep 2025 – Present
Orsay, France

Magistère in Computer Science (Honors Research Program)
University Paris-Saclay
Graduated with Honors

Sep 2024 – Aug 2025
Orsay, France

Double Bachelor's Degree in Computer Science and Mathematics
University Paris-Saclay

Sep 2022 – Aug 2024
Orsay, France

Intensive English Language Program
EF New York
Achieved C2 Proficiency

Sep 2019 – May 2020
New York, USA

PUBLICATIONS

Revisiting Class Imbalance Ratios in Imbalanced Learning
Junior Conference on Data Science & Engineering, 2025

Fine-Grained Mention-Level Analysis of Biomedical Entity Linking Models
Medical Informatics Europe, 2026

PROFESSIONAL EXPERIENCE

Supervised Research Project
LISN - Laboratoire Interdisciplinaire des Sciences du Numérique

Jan 2026 – Mar 2026
Orsay, France

- Investigated Biomedical Entity Linking (BEL) strategies to normalize PICO entities (Population, Intervention, Comparison, Outcome) to the MeSH knowledge base. Developed and evaluated a hybrid normalization pipeline, comparing a custom rule-based script (heuristics) with ArboEL (a state-of-the-art graph-based entity linking model).

Contact: Nona Naderi (LISN, Université Paris-Saclay)

Research Internship
LISN - Laboratoire Interdisciplinaire des Sciences du Numérique

May 2025 – Aug 2025
Orsay, France

- Analyzed biomedical entity linking models on the BELB benchmark, focusing on generalization to rare or complex mentions. Developed quantitative and visual analyses of dataset characteristics (mention length, ambiguity, frequency) and their impact on prediction quality. Compared recent models, identified consistent weaknesses, and proposed improvements; published a paper in MIE

Contact: Nona Naderi (LISN, Université Paris-Saclay)

Supervised Research Project
LISN - Laboratoire Interdisciplinaire des Sciences du Numérique

Jan 2025 – May 2025
Orsay, France

- Studied the impact of class imbalance on classification tasks using a spherical Teacher-Student perceptron. Conducted experiments in Python (Scikit-Learn, NumPy, Matplotlib) with different noise levels, loss functions, and training methods (GD, Langevin dynamics). Demonstrated that the optimal imbalance ratio in training sets differs from 0.5 in class-imbalanced problems.

Contact: François Landes (LISN, Université Paris-Saclay)

Generative AI Trainer
Outlier / Alignerr

Jan 2025 – Present
Remote

- Designed and refined prompts to enhance the performance of generative AI models. Reviewed and corrected AI-generated outputs to ensure accuracy and quality. Contributed to the continuous improvement of deep learning models through feedback and prompt optimization.

PROJECTS

Structure Detection in Fusion Plasma Simulations ⊘
- Developed a multi-stage detection pipeline featuring a YOLOv8 architecture optimized with a custom MLP post-filter for low-contrast frames, incorporating a pseudo-labeling strategy to expand a limited 30-image dataset and a feature extractor targeting intensity statistics and Sobel gradients.

Measuring Market Impact of Financial News ⊘
- Engineered a computationally efficient framework to transform unstructured financial news into (date, ticker, impact summary) event representations by implementing a hierarchical Map-Reduce summarization architecture with compact Transformers, ensuring numerical fidelity and entity grounding while validating results through a finance-oriented LLM-as-a-judge audit.

NBA MVP Prediction Model ⊘
- Collected and cleaned player and team statistics from publicly available online sources to build a structured dataset. Performed feature selection and engineering, experimenting with different attribute selection strategies. Using several predictive models implemented with Scikit-Learn, achieved an accuracy >80% in predicting the actual MVP outcome.

Traffic Sign Recognition Model ⊘
- Developed a machine learning model to recognize traffic signs from pictures. Implemented preprocessing (normalization, resizing, feature extraction) and trained multiple supervised learning algorithms, reaching over 95% accuracy.

Dual Sudoku AI Agent ⊘
- Designed an artificial intelligence to solve and play Dual Sudoku by combining search algorithms and heuristic strategies. Implemented efficient state representation and evaluation functions to handle the game's combinatorial complexity.

Java-like Interpreter ⊘
- Built an interpreter supporting basic arithmetic, classes, methods, and Java-like type checking. Entirely coded in OCaml with Ocamllex and Menhir.

Other projects
- Developed various algorithmic implementations (e.g., Graphs, Linked Trees, Quadtrees) and 2D games (e.g., Air Hockey, Labyrinth solver, Colt Express, Worms) across Python, OCaml, Java, and C++.

SKILLS

Programming:
- Python (PyTorch, Scikit-Learn, NumPy, Matplotlib), Bash, C/C++, Java, OCaml, SQL

Artificial Intelligence:
- Deep Learning, Natural Language Processing, Computer Vision

Tools:
- LaTeX, Git, Linux, Slurm (Cluster Computing), Microsoft Office Suite

ACHIEVEMENTS

Winner of an AI competition by designing the best-performing Dual Sudoku agent

3x Prologin finalist, a national algorithmic programming contest

LANGUAGES

French
Native

Russian
Conversational

English
Bilingual Proficiency
TOEFL iBT: 108/120
TOEIC: 990/990