Davide Baldelli

✉ davide.baldelli98@gmail.com | 🏠 dundalia.github.io | github.com/Dundalia | linkedin.com/in/davide-baldelli-b55618203/

Education

MILA and Polytechnique Montréal *Montréal, Canada*
PhD in Computer Engineering *Jan 2025 - Current*

- GPA: 4.0
- **Supervisors:** Sarath Chandar, Amal Zouaq.
- **Courses:** Representation Learning, Reinforcement Learning.

University of Bologna *Bologna, Italy*
MSc in Artificial Intelligence *Sept 2021 - Oct 2023*

- Grade: 110/110 cum laude.
- **Courses:** Programming for Data Science, Knowledge Representation and Reasoning, Statistical Theory and Methods, Machine Learning, Deep Learning, Computer Vision, Natural Language Processing.

University of Florence *Florence, Italy*
BSc in Mathematics *Sept 2017 - Mar 2021*

- Grade: 110/110.
- **Courses:** Mathematical Analysis, Linear Algebra, Group Theory, Geometry, Probability and Statistics, Logic and Computability.

Research Experience

National Institute of Informatics *Tokyo, Japan*
MSc Thesis - International Internship Program [**Optional - Own Initiative**] *Mar 2023 - Aug 2023*

- **Type**: Optional research internship undertaken on own initiative (not mandatory curriculum requirement)
- **Supervisors**: Akiko Aizawa (National Institute of Informatics), Junfeng Jiang (National Institute of Informatics), Paolo Torroni (University of Bologna)
- **Title**: A TWO-step LLM Augmented distillation method for passage Reranking
- Conducted research on Text Information Retrieval. We have investigated recent advancements in open-sourced Large Language Models, text retrievers, rerankers, knowledge distillation and ranking losses. Our approach is to adapt a T5 model to solve the reranking task, training it on GPT augmented labels. We have obtained SoTA results on the TREC-DL test sets and on the BEIR benchmark.

University of Florence *Florence, Italy*
BSc Thesis [**Mandatory Curriculum**] *Sept 2020 - Mar 2021*

- **Type**: Mandatory research training as part of university curriculum
- **Supervisor**: Francesca Romana Nardi (University of Florence)
- **Title** (translated): Big Deviations for Markov Chains
- The thesis focused on presenting a theorem describing the asymptotic behavior of Markov chains for studying rare events within the context of large deviations theory.

Publications

PEER-REVIEWED CONFERENCE PROCEEDINGS

Baldelli, Davide, Junfeng Jiang, Akiko Aizawa, and Paolo Torroni (2024). "TWOLAR: A TWO-Step LLM-Augmented Distillation Method for Passage Reranking". *Advances in Information Retrieval*. Vol. 14608. Lecture Notes in Computer Science. Springer, pp. 465–481. DOI: `10.1007/978-3-031-56027-9_29`. URL: `https://doi.org/10.1007/978-3-031-56027-9_29`.

PRE-PRINTS AND WORKING PAPERS

Govindarajan, Prashant, **Baldelli, Davide**, Jay Pathak, Quentin Fournier, and Sarath Chandar (2025). *CADmium: Fine-Tuning Code Language Models for Text-Driven Sequential CAD Design*. arXiv: 2507.09792. URL: `https://arxiv.org/abs/2507.09792`.

Conferences & Presentations

46th European Conference on Information Retrieval (ECIR 2024) *Glasgow, UK*
Oral Presentation *Mar 2024*

- **Paper:** TWOLAR: A TWO-step LLM-Augmented Distillation Method for Passage Reranking
- Presented research on text information retrieval and large language models for passage reranking.

Professional Experience

MILA and Polytechnique Montréal *Montréal, Canada*

Teaching Assistant - Summer High School Internships *Summer 2025*

- Designed and delivered educational workshops on Machine Learning fundamentals and Probability & Statistics for high school students.
- Mentored students in hands-on learning activities, fostering interest in AI research.
- Contributed to science outreach by making advanced technical concepts accessible to younger audiences and encouraging future academic pursuit in AI.

Independent Consultant (Upwork Platform) *Remote*

Freelance Machine Learning Engineer *June 2024 - Jan 2025*

- Delivered a variety of machine learning projects for international clients, consistently earning 5-star ratings for quality, speed, and communication.

Loop AI Group *Full Time - Fully Remote*

Machine Learning Engineer & Master's Thesis Co-Supervisor *Sept 2023 - Dec 2024*

- Co-supervised a master's thesis titled "Self Assessment Tool for Medical Exam Candidates" in collaboration with the University of Padova. Provided mentorship throughout the entire research process, from problem formulation to final defense. Guided the candidate in literature review, experimental design, machine learning model selection, and evaluation methodology.
- Conducted deep learning project for musical Deepfake detection algorithm achieving 94% of F1 score on a proprietary italian deepfake music dataset.
- Conducted deep learning project for automatic translation of audio data for content creation.
- Implemented statistical and deep learning based time series analysis models to solve demand forecasting.
- Implemented BERT-based email classifier tool for an italian University's administrative office, achieving 98% accuracy.

Loop AI Group *Part Time - Fully Remote*

Data Scientist Consultant *Sept 2022 - Sept 2023*

- Implemented closed-domain generative question-answering systems on technical topics, exploiting SoTA Pretrained Seq2Seq Language Models.
- Implemented rule-based and transformers-based text retrievers.

Leadership & Community Involvement

Belle Parole Arci APS *Florence, Italy*

Leadership in Non-Profit Organization *Nov 2023 - Present*

- Involved in founding and leading Belle Parole Arci APS, a non-profit organization.
- Responsible for organizing community events and managing bar operations as part of the organization's activities.
- Emphasized community engagement, cultural enrichment, and team coordination in a non-profit setting.
- Built the website of the organization: **belleparoleaps.it**.

Other Achievements

AWARDS & RECOGNITIONS

Jun 2025 **Selected Participant**, Y Combinator AI Startup School *San Francisco, CA*

Jun 2022 **4th place**, Loop Q Prize Competition *Italy*

CERTIFICATIONS

Apr 2025 **Agents Course**, Hugging Face *Remote*

Skills

Programming	Python (fluent); HTML, CSS, R, MatLab, Minizinc, NetLogo (familiar).
Data Science	Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly
Deep Learning	Pytorch, Pytorch-lightning, Tensorflow.
NLP	HuggingFace, Transformers, Prompt Engineer, OpenAI GPT.
Miscellaneous	Linux, Shell, LaTeX(Overleaf/R Markdown), Git, Docker, Figma, Blender.

Languages

English	Professional proficiency
Italian	Native proficiency

References available upon request.