# Davide Baldelli

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

## Personal Profile

PhD Student at MILA and Polytechnique Montréal.

## 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.

**Liceo Scientifico Guido Castelnuovo**                                  *Florence, Italy*

Scientific High School                                                   *Sept 2012 - June 2017*

- Passed with 96/100.

## Publications

### CONFERENCE PROCEEDINGS

TWOLAR: a TWO-step LLM-Augmented distillation method for passage Reranking
  Davide Baldelli, Junfeng Jiang, Akiko Aizawa, Paolo Torroni
  *Proceedings of the European Conference on Information Retrieval (ECIR 2024)*, 2024

### PRE-PRINTS

CADmium: Fine-Tuning Code Language Models for Text-Driven Sequential CAD Design
  Prashant Govindarajan, Davide Baldelli, Jay Pathak, Quentin Fournier, Sarath Chandar
  *arXiv preprint arXiv:2507.09792 (2025). 2025*

## Achievements

Jun 2025 **Selected Participant**, Y Combinator AI Startup School                 *San Francisco, CA*
Jun 2022 **4th place**, Loop Q Prize Competition                                   *Italy*

## Research Experience

**National Institute of Informatics**                                    *Tokyo, Japan*

MSc Thesis - International Internship Program                            *Mar 2023 - Aug 2023*

- **Advisors**: Akiko Aizawa, Junfeng Jiang, Paolo Torroni
- **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*

<u>BSc Thesis</u>                                                                          *Sept 2020 - Mar 2021*

- **Advisor**: Francesca Romana Nardi
- **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.

## Work Experience

**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                                                                   *Sept 2023 - Dec 2024*

- Co-supervised a master's thesis titled <u>"Self Assessment Tool for Medical Exam Candidates"</u> in collaboration with the University of Padova. Guided the candidate in applying machine learning techniques to develop a self-assessment tool. Assisted in research direction, methodology selection, and technical problem-solving.
- 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.

## Certifications

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

## 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**.

## 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. |
| **Web Development** | Flask, Webflow |
| **Miscellaneous** | Linux, Shell, LaTeX(Overleaf/R Markdown), Git, Docker, Figma, Blender. |
| **Soft Skills** | Time Management, Teamwork, Problem-solving, Documentation, Engaging Presentation. |

## Languages

| | |
|---:|---|
| **English** | Professional proficiency |
| **Italian** | Native proficiency |

**References available upon request.**