David La Barbera

Via Sottomonte 61 • Venzone (UD), 33010, Italy • labarberadavid92@gmail.com • +39 346 4317250

Summary Statement

I work on system-level AI pipelines combining retrieval, LLMs, and human-in-the-loop components, with a strong focus on evaluation, failure analysis, and reliability under real-world constraints.
Looking for EU-friendly remote roles where applied research directly informs production-oriented AI systems.

Work Experience

Contract Professor — Pavia, Italy
University of Pavia — A.Y. 2025/2026
Course: Information Retrieval and Recommender Systems (6 CFU), BSc in Artificial Intelligence

Post-Doctoral Researcher — Milan, Italy
University of Milano-Bicocca — 01/02/2025 – Present
Activities: Developed an LLM-driven agent combining structured persona representations, retrieval-augmented generation (RAG), session memory, and safety constraints to simulate realistic therapeutic dialogues and emotional trajectories. Built experimental setups to analyze agent behavior, reliability, and failure modes in sensitive educational decision-support scenarios, with particular attention to interpretability and risk mitigation in human-facing interactions. Performed extensive testing with human feedback.

Research Fellow — Udine, Italy
Central Friuli University Health Authority — 01/01/2021 – 31/10/2021
"Audit and Feedback in Emergency: Data Analytics on Territorial Relief Times"
Activities: data analysis and modeling of emergency response times at the territorial level. Designed and implemented a software simulator to explore alternative resource allocation and response-time scenarios, enabling what-if analyses to support decision-making and service optimization. Evaluated system behavior under varying operational assumptions.

Research Fellow — Udine, Italy
University of Udine — 01/12/2019 – 30/11/2020
Project title: "Artificial Intelligence for the Diagnosis of Breast Cancer on Microscopy Images"
Activities: development and evaluation of AI methods for medical image analysis, focusing on breast cancer diagnosis from histopathological slides. Designed deep learning pipelines using multi-stage and multiple instance learning approaches, and conducted benchmarking experiments on international datasets to analyze performance, robustness, and failure cases.

Full Stack Developer — Udine, Italy
DataMantix Srl — 01/10/2017 - 31/12/2017
Activities: freelance collaboration for development of an ecommerce platform.

Education

Ph.D. in Computer Science and Artificial Intelligence — Udine, Italy
University of Udine, XXXVII Cycle — 01/11/2021 – 26/03/2025
Thesis: Human-AI Collaboration in Fact-Checking
Activities: research on misinformation detection through Human-In-The-Loop systems. Developed AI methods to

process large-scale online data while integrating human intelligence via crowdsourcing to perform complex tasks, improve reliability, and enhance explainability. Investigated strategies for fast identification of false information and its impact on human decision-making in online communication.
Grade: **Excellent cum Laude**

MS in Computer Science
Udine, Italy

University of Udine   21/10/2016 – 18/10/2019

Thesis: "Fake News Detection through Crowdsourcing: Use of Multi-Valued Scales and Analysis of Political Bias"
Grade: **110 cum Laude**/110

BS in Web and Multimedia Technologies
Udine, Italy

University of Udine   01/09/2012 – 20/10/2016

Thesis: "Use of Semiotic Tools for the Evaluation of Commercial Videos in Brand Communication"
Supervisor: Prof. Elio Toppano
Grade: **96**/110

Diploma in Industrial Chief Technician with IT Specialization
Gemona Del Friuli, Italy

I.S.I.S. R. D'Aronco   01/09/2006 - 30/06/2011

Grade: **70**/110

Skills

High-Level Skills:
- End-to-end ML system design for decision support
- Retrieval-Augmented Generation (RAG) and information access pipelines
- Human-in-the-loop system design and evaluation
- Experimental design, error analysis, evaluation metrics,
- Data analytics and visualization
- Scientific and technical communication

Technical Skills:
- **Core:** Python, Pandas, SciPy, Matplotlib, Seaborn
- **ML/NLP:** PyTorch, Hugging Face Transformers, vector embeddings, vLLM
- **Systems & Prototyping:** LangChain, FastAPI, Flask, Dash
- **Data & Infra:** SQL / NoSQL, schema design, data quality validation
- **Tooling:** Git, Docker, Linux shell, Jupyter, VS Code, LaTeX

Languages:
- **Italian:** Native speaker; **Friulano:** Native speaker; **English:** C1 level

Awards

- **Best Paper Award** at "ICTIR 2025", for the paper "Impersonating the Crowd: Evaluating LLMs' Ability to Replicate Human Judgment in Misinformation Assessment".
- **Best Paper Award** at "NL4AI 2022", for the paper "A Hybrid Human-In-The-Loop Framework for Fact Checking".
- **Best Paper Award** at "ECIR 2020", for the paper "Crowdsourcing Truthfulness: The Impact of Judgment Scale and Assessor Bias".
- **2nd Place at the HEROHE challange** for the development of a cascade deep learning classifiers via Multi-Instance Learning for detecting HER2 from Haematoxylin-Eosin Slides.

Selected Research Publications

5. *HEROHE Challenge: Predicting HER2 Status in Breast Cancer from Hematoxylin–Eosin Whole-Slide Imaging*, Eduardo Conde-Sousa, João Vale, Ming Feng, Kele Xu, Yin Wang, Vincenzo Della Mea, **David La Barbera**, Ehsan Montahaei, Mahdieh Baghshah, Andreas Turzynski, Jacob Gildenblat, Eldad Klaiman, Yiyu Hong, Guilherme Aresta, Teresa Araújo, Paulo Aguiar, Catarina Eloy, Antonio Polónia. *Journal of Imaging*, 8(8), 2022, Article 213. DOI: 10.3390/jimaging8080213.

4. *Impersonating the Crowd: Evaluating LLMs' Ability to Replicate Human Judgment in Misinformation Assessment*, **David La Barbera**, Riccardo Lunardi, Mengdie Zhuang, Kevin Roitero. *Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)*, DOI: 10.1145/3731120.3744581.

3. *A Comparative Analysis of Retrieval-Augmented Generation and Crowdsourcing for Fact-Checking*, Francesco Bombassei De Bona, **David La Barbera**, Stefano Mizzaro, Kevin Roitero. *Proceedings of the 47th European Conference on Information Retrieval (ECIR 2025)*, pp. 446–454. DOI: 10.1007/978-3-031-88714-7_44.

2. *The Elusiveness of Detecting Political Bias in Language Models*, Riccardo Lunardi, **David La Barbera**, Kevin Roitero. *Proceedings of the 33rd Conference on Information and Knowledge Management (CIKM 2024)*, pp. 3922–3926. DOI: 10.1145/3627673.3680002.

1. *Combining Large Language Models and Crowdsourcing for Hybrid Human-AI Misinformation Detection*, Xia Zeng, **David La Barbera**, Kevin Roitero, Arkaitz Zubiaga, Stefano Mizzaro. *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2024)*, pp. 2332–2336. DOI: 10.1145/3626772.3657965.