Yannis Karmim

Paris – Santiago yanniskarmim@gmail.com +33 6 52 08 07 85 ykrmm.github.io

Yannis Karmim ykrmm Scholar

AI researcher, PhD 2025. Five first-author papers at **NeurIPS**, **ICCV**, **TMLR**, **EMNLP** and **EACL**, spanning graph representation learning, uncertainty estimation LLM reasoning and ent-to-end LM training. Hands-on with large-scale distributed training (multi-node multi-GPU) and with deployment under tight compute constraints. Seeking a research scientist role where rigorous empirical methodology meets large-scale systems.

Research

Postdoctoral Researcher, INRIA Paris & INRIA Chile 2025–

Small reasoning models and agents, knowledge acquisition, knowledge composition. Socio-cultural bias evaluation (EACL'26) and multi-hop reasoning via pretraining augmentation (EMNLP'26).

Expertise: Safety (ICCV'25), LLMs (EACL'26, EMNLP'26), Reasoning (EMNLP'26), Agents (ICMLC'26), Graphs (NeurIPS'24, ECML'24, TMLR).

PhD in Machine Learning, CNAM Paris 2021–2025

Machine learning on graphs for recommendation and dynamic prediction. Defended June 3, 2025.
Supervisors: Nicolas Thome, Raphaël Fournier S'niehotta. **Labs:** CEDRIC (CNAM); ISIR (Sorbonne University).
Areas: Graph Neural Networks, Temporal Data, Uncertainty, Transformers, Vision-Language Models, LLMs.

Selected publications

- Poulenard P., **Karmim Y.**, Barrière V. **Knowledge-Graph Based Augmentation versus Retrieval Augmented Generation for Cultural-Related Question Answering under review at EMNLP'W 2026**.
- **Karmim Y.**, Marti L., Seddah D., Barrière V. **Multi-Hop Knowledge Composition is Bound by Pretraining Exposure Findings EMNLP 2026**.
- **Karmim Y.**, Pino R., Contreras H., Lira H., Cifuentes S., Escoffier S., Martí L., Seddah D., Barrière V. **Leveraging Wikidata for Geographically Informed Sociocultural Bias Dataset Creation: Application to Latin America EACL'W 2026**.
- Hilel M., **Karmim Y.**, De Bodinat J., Sarehane R., Gillon A. **Building Specialized Software-Assistant ChatBot with Graph-Based Retrieval-Augmented Generation ICMLC 2026**.
- **Karmim Y.**, Lafon M., Rodriguez J., Couairon P., Rambour C., Fournier R., Ben-Hayed I., Dolz J., Thome N. **Learning Vision-Language Uncertainties for Failure Prediction. ICCV 2025**.
- **Karmim Y.**, Lafon M., Fournier R., Thome N. **Supra-Laplacian Encoding for Transformer on Dynamic Graphs. NeurIPS 2024**.
- **Karmim Y.**, Yang L., Fournier R., Chatelain C., Adam S., Thome N. **Temporal Receptive Field in Dynamic Graph Learning: A Comprehensive Analysis. ECML'W 2024**.
- **Karmim Y.**, Ramzi E., Fournier R., Thome N. **ITEM: Improving Training and Evaluation of Message-Passing based GNNs for Top-k Recommendation. TMLR 2023**.

Industry & engineering experience

R&D Consultant in Deep Learning *Paris, France*
RAKAM May 2024 – Sept 2024

- Built an automatic knowledge-graph construction pipeline from large-scale structured data
- RAG-based LLM conditioning for question answering over the generated graph, 30-70% gain over baseline .
- Fast graph partitioning and inference for real-time tutorial generation.

Research Engineer *Paris, France*
Conservatoire National des Arts et Métiers Sept 2020 – Oct 2021

- Spatio-temporal object detection for real-time defense systems (aerial drone detection and localisation).
- Lightweight model trained by large-scale pre-training (1M) and fine-tuned semi-supervised on few labelled (1k) examples.
- 95% mAP under test-time and deployment constraints.
- Optimised inference on NVIDIA Jetson Nano
- *Deeplomatics: A Deep-Learning Based Multimodal Approach for Aerial Drone Detection and Localization.* **QUIET DRONES 2022**.

Technical expertise

Languages: Python
ML stack: PyTorch, PyTorch Lightning, PyTorch Geometric, HuggingFace (Transformers, Datasets, Accelerate), NumPy, SciPy, pandas, scikit-learn, Weights & Biases, TransformerLens (interpretability)
Scale & systems: Distributed Data Parallel (FSDP / DeepSpeed), multi-node multi-GPU training up to 100B tokens, 400M images (LAION), SLURM, CUDA, mixed precision, profiling and memory optimisation, Docker, Git
Domains: Graph neural networks, large language models, vision-language models, uncertainty quantification and calibration, edge deployment (NVIDIA Jetson)
Spoken languages: French (native), English (C1), Portuguese (B1), German (B1)

Teaching, service & talks

Teaching 2021–2026 — 140 hours. Teaching Assistant, CNAM & Sorbonne Université (2021–2024); Instructor, Jedha (2025–2026).

- *Graduate:* Probability, Statistics and Learning · Reinforcement Learning · Machine Learning and Python · Deep Learning · Graph Learning
- *Jedha:* Generative AI · MLOps · Workflow Orchestration · Reinforcement Learning

Reviewer: NeurIPS, ICML, ICLR, CVPR, EMNLP, TMLR, ICLR Workshops

Invited talks: CENIA, Santiago (May 2026) · Universidad de Chile (May 2026) · INRIA Chile (Mar. 2026) — multi-hop reasoning and knowledge acquisition in LLMs.
Conference presentations: NeurIPS 2024, Vancouver (poster, *Supra-Laplacian Encoding*) · ECML Graph Learning Workshop, Vilnius 2024 (poster, *Temporal Receptive Field*).

Summer school: Peyresq Summer School on Graph Signal Processing (2023)

Education

PhD in Machine Learning — Conservatoire National des Arts et Métiers, Paris — Oct 2021 – June 2025
MSc in Data Science — Sorbonne Université, Paris — Sept 2018 – June 2020
Exchange programme, Computer Science — Université de Montréal, Canada — Jan 2018 – Jul 2018
BSc in Computer Science — Sorbonne Université, Paris — Sept 2015 – Jan 2018