Alexandre **Ramé**

Research scientist at Google DeepMind. PhD in deep learning at Sorbonne Université.
Investigating how model merging, RL and the generalization literature can align AIs with the world *in all its diversity*.

Paris, France

+33 6.32.17.09.14 | alexandre.rame.cl@gmail.com | alexrame.github.io | alexrame

Education

PhD in Computer Science and Deep Learning — *Paris, France*

Sorbonne University (ISIR, MLIA) *Advisor: Pr. Matthieu Cord* — *Mar 2020 - Oct 2023*

- Manuscript: *Diverse and Efficient Ensembling of Deep Networks*.
- Topics: model merging, weight averaging, robustness, out-of-distribution generalization, continual learning and alignment.
- Received the award of the best French PhD from SSFAM.

Master of Science in Operations Research GPA: 3.9 / 4.0 — *New York, USA*

Columbia University — *Sep 2014 - May 2015*

- Majors: optimization and machine learning.
- Minors: deep learning, statistics and programming.

Diplôme d'Ingénieur Polytechnicien GPA: 3.7 / 4.0 — *Palaiseau, France*

Ecole Polytechnique — *Sep 2011 - May 2014*

- Major in applied mathematics: optimization, probability, statistics, stochastic finance and times series analysis.
- Minors: computer science, economics, physics, entrepreneurship and mathematics.

MPSI - MP* Info — *Versailles, France*

Lycée Sainte-Geneviève — *Sep 2009 - Jul 2011*

Mathematics, physics and computer science.

Experience

Google DeepMind — *Paris, France*

Research scientist *Advisor: Dr. Olivier Bachem* — *Mar 2023 -*

- RLHF Alignment of Gemini LMMs for improved quality and safety.
- Reward modeling and reinforcement learning.

Google DeepMind — *Paris, France*

Student Researcher *Advisor: Dr. Johan Ferret* — *Oct 2023 - Jan 2024*

- Improving the robustness of reward models for RLHF.

FAIR Meta AI — *Paris, France*

Research Scientist Intern in the Fairness and Robustness team *Advisor: Dr. David Lopez-Paz and Dr. Léon Bottou* — *Sep 2022 - Feb 2023*

- Investigating how weight averaging strategies can improve out-of-distribution generalization.
- Exploring how the updatable machine learning paradigm can help for embarrassingly simple parallelization of large-scale trainings.

Heuritech — *Paris, France*

Research Scientist in Deep Learning *Advisor: Dr. Charles Ollion* — *Jan 2016 - Nov 2019*

- Main contributor of the computer vision pipeline. Implementing and improving deep classification and detection models.

Flaminem — *Paris, France*

Research Scientist in Machine Learning — *Sep 2015 - Dec 2015*

- Big data challenges to predict long-term purchase decision.

Selected Publications

WARM: On the Benefits of Weight Averaged Reward Models — *ICML*
Alexandre Ramé, Nino Vieillard, Léonard Hussenot, Robert Dadashi, Geoffrey Cideron, Olivier Bachem, Johan Ferret — *2024*

Direct Language Model Alignment from Online AI Feedback — *arXiv*
Shangmin Guo, Biao Zhang, Tianlin Liu, Tianqi Liu, Misha Khalman, Felipe Llinares, Alexandre Ramé, Thomas Mesnard, Yao Zhao, Bilal Piot, Johan Ferret, Mathieu Blondel — *2024*

Beyond task performance: Evaluating and reducing the limitations of large multimodal models with in-context-learning? — *ICLR*
Mustafa Shukor, Alexandre Ramé, Corentin Dancette, Matthieu Cord — *2024*

Rewarded Soups: Towards Pareto-Optimal Alignment by Interpolating Weights Fine-tuned on Diverse Rewards — *NeurIPS*
Alexandre Ramé, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, Matthieu Cord — *2023*

UniVAL: Unified Model for Image, Video, Audio and Language Tasks — *TMLR*
Mustafa Shukor, Corentin Dancette, Alexandre Ramé, Matthieu Cord — *2023*

Model Ratatouille: Recycling Diverse Models for Out-of-Distribution Generalization — *ICML*
Alexandre Ramé, Kartik Ahuja, Jianyu Zhang, Matthieu Cord, Léon Bottou, David Lopez-Paz — *2023*

Diverse Weight Averaging for Out-of-Distribution Generalization — *NeurIPS*
Alexandre Ramé, Matthieu Kirchmeyer, Thibaud Rahier, Alain Rakotomamonjy, Patrick Gallinari, Matthieu Cord — *2022*

DyTox: Transformers for Continual Learning with DYnamic TOken eXpansion — *CVPR*
Arthur Douillard, Alexandre Ramé, Guillaume Couairon, Matthieu Cord — *2022*

Fishr: Invariant Gradient Variances for Out-of-distribution Generalization — *ICML*
Alexandre Ramé, Corentin Dancette, Matthieu Cord — *2022*

MixMo: Mixing Multiple Inputs for Multiple Outputs via Deep Subnetworks — *ICCV*
Alexandre Ramé, Remy Sun, Matthieu Cord — *2021*

DICE: Diversity in Deep Ensembles via Conditional Redundancy Adversarial Estimation — *ICLR*
Alexandre Ramé, Matthieu Cord — *2021*

OMNIA Faster R-CNN: Detection in the Wild through Dataset Merging and Soft Distillation — *arXiv*
Alexandre Ramé, Emilien Garreau, Hedi Ben-Younes, Charles Ollion — *2018*

Leveraging Weakly Annotated Data for Fashion Image Retrieval and Label Prediction — *ICCVW*
Charles Corbiere, Hedi Ben-Younes, Alexandre Ramé, Charles Ollion — *2017*

Teaching

Teacher Assistant
Sorbonne Université · *Master level* · Deep Learning for Computer Vision — *Fall 2020 / Fall 2021*

Teacher Assistant
Data Science l'X-Paris Saclay · *Master level* · Deep Learning — *Fall 2017 / Fall 2018*

Volunteer Teacher and Youth Leader

Fondation d'Auteuil Sannois · Mathematics *Nov 2011 - Mar 2012*

Skills

- **Programming Languages**: Python · Shell · Scala · R
- **Packages**: PyTorch · JAX · Tensorflow / Keras · Theano · Scikit-Learn · Numpy · Pandas
- **Tools & OS**: Linux · Latex · Git · Jupyter/Colab · Vim · VSCode
- **Languages**: French (native) · English (fluent) · Spanish (beginner)
- **Reviewing**: NeurIPS (top reviewer 2023) · ICML · ICLR · CVPR · CoLLAs · IJCV

Main Talks

Sorbonne ISIR, Paris Weight averaged reward models *Jan 2024*

Google DeepMind, Paris Efficient, reliable and robust reward models with weight averaging *Dec 2023*

ENPC Imagine, Paris Diverse and efficient ensembling of deep networks *Nov 2023*

INRIA Sierra, Paris Diverse and efficient ensembling of deep networks *Nov 2023*

Valeo.ai, Paris Diverse and efficient ensembling of deep networks *Sept 2023*

INRIA THOTH, Grenoble Weight Averaging for Generalization and Alignment *July 2023*

Samsung SAIL, Montréal (Canada) Weight Averaging and Diversity for Generalization *June 2023*

ECML KDD, Grenoble A Bias-Variance Analysis of Out-of-Distribution Generalization *Sep 2022*

Facebook AI Research, Paris Fishr for Domain Generalization *Oct 2021*

Valeo.ai, Paris Dice for Diversity in Deep Ensembles *Mar 2021*

Paris Deep Learning Meetup #16, Paris OMNIA Faster R-CNN for Semi-Supervised Object Detection *Jan 2019*

Paris Deep Learning Meetup #6, Paris Correlational Neural Networks for Multilingual Embeddings *Feb 2017*