# Hamza Cherkaoui

hamza.cherkao@gmail.com
hcherkaoui.github.io  Hamza Cherkaoui (h 8 – i10 8)  hcherkaoui  Hamza Cherkaoui

## Research Profile

Machine learning researcher working on: *generative modeling, conditional flow matching, diffusion models and transfer learning.*

My recent work studies heavy-tailed generative modeling and the training dynamics of diffusion and flow-based models. Concretely, I combine mathematical analysis with PyTorch-based large-scale empirical evaluation to develop and assess diffusion and flow-based generative methods.

## Research Experience

**2025–present** — **Postdoctoral Researcher, Télécom SudParis**, Palaiseau, France
Generative models

- *Collaborator*: Prof. Hélène Halconruy.
- Developed theoretical analyses and empirical benchmarks for diffusion and flow-matching models under heavy-tailed data distributions.
- Studied the trade-off between initialization, training error, tail coverage, and sample diversity in generative modeling.
- Built PyTorch/SLURM research pipelines for reproducible large-scale benchmarks.
- Two first-author preprints under review.

**2022–2025** **Postdoctoral Researcher, Huawei Technologies**, Paris, France
Bandits / transfer

- *Collaborator*: Dr. Igor Colin.
- Developed algorithms for multi-agent linear bandits, adaptive sample sharing, transfer learning, and safe best-arm identification.
- Published two first-author conference papers, including ICML 2025, and wrote two first-author preprints under review.

**2021** — **Postdoctoral Researcher, BioMaps research team**, CEA-Saclay, Orsay, France
Neuroimaging

- *Collaborators*: Dr. Claire Leroy and Dr. Nicolas Tournier.
- Studied the estimation of the haemodynamic response function from fMRI data in a pharmacological context.
- Published one first-author international conference paper.

**2017–2021** **PhD, PARIETAL research team**, CEA-Saclay / INRIA-Saclay, France
PhD program

- *Supervisors*: Dr. Philippe Ciuciu, Dr. Thomas Moreau, and Dr. Claire Leroy.
- *Thesis*: Efficient whole-brain estimation of the haemodynamic response function for TV-regularized semi-blind deconvolution of neural activity in fMRI.
- Published five first-author international conference papers, including NeurIPS 2020, and one first-author journal article.

## Selected Publications

[1] **Cherkaoui, H.**, Halconruy, H., and Ocello, A. *"Do Heavy Tails Help Diffusion? On the Subtle Trade-off Between Initialization and Training"*, preprint under review.

[2] **Cherkaoui, H.**, Halconruy, H., and Petetin, Y. *"When to Transfer: Adaptive Source Selection for Positive Transfer in Linear Models"*, preprint under review.

[3] **Cherkaoui, H.**, Tiomoko, M., Paolo, G., Zhang, Y., Meng, Y., Keli, Z., and Tiomoko Ali, A. *"Post-Training Corrections for Improved Time-Series Forecasting"*, preprint under review.

[4] **Cherkaoui, H.**, Colin, I., Shang, X., and Barlier, M. *"Price of Safety in Linear Best Arm Identification"*, preprint under review.

[5] Tiomoko, M., **Cherkaoui, H.**, Seddik, M., Louart, C., Schnoor, E., and Kegl, B. *"High-Dimensional Analysis of Bootstrap Ensemble Classifiers"*, AISTATS, 2026.

[6] **Cherkaoui, H.**, Barlier, M., and Colin, I. *"Adaptive Sample Sharing for Multi-Agent Linear Bandits"*, ICML, 2025.

[7] **Cherkaoui, H.**, Moreau, T., Halimi, A., Leroy, C., and Ciuciu, P. *"Multivariate Semi-Blind Deconvolution of fMRI Time Series"*, NeuroImage, 2021.

[8] **Cherkaoui, H.**, Sulam, J., and Moreau, T. *"Learning to Solve TV-Regularized Problems with Unrolled Algorithms"*, NeurIPS, 2020.

[9] **Cherkaoui, H.**, Moreau, T., Halimi, A., and Ciuciu, P. *"Sparsity-Based Blind Deconvolution of Neural Activation Signal in fMRI"*, ICASSP, 2019.

## Research Software

**2026** **FlowBench**, PyTorch benchmark for diffusion and flow-matching models.
- Designed for reproducible comparison of Gaussian and heavy-tailed generative models across synthetic, tabular, and image-like datasets.

**2025** **Metis**, AutoML toolbox for model selection and ensembling.
- Designed for automated model selection, ensembling, and predictive-performance optimization.

**2023** **Bandpy**, Gym-compatible benchmark for single- and multi-agent bandit algorithms.
- Designed to streamline reproducible bandit experiments and algorithmic comparisons.

**2020** **Carpet**, fast minimization of total-variation regularized problems. carpet
- Provides tools for one-dimensional TV-regularized inverse problems and unrolled neural solvers.

**2019** **HemoLearn**, multivariate estimation of the HRF from fMRI data. hemolearn
- Disentangles neurovascular coupling from neural activation signals in fMRI.

## Languages & Technical Skills

- *Languages:* French (native) English (TOEIC 970/990, fluent)
- *Programming:* `Python`; working knowledge of `Cython`, `C`, and `C++`
- *Deep learning stack:* `PyTorch`, `JAX/Flax`, `PyTorch Geometric`, `NumPy`, `SciPy`, `Numba`, `scikit-learn`
- *Research engineering:* large-scale ML experiments, reproducible pipelines, GPU/HPC training
- *Tools:* `Git`, LaTeX, Linux/HPC environments, SLURM

## Reviewing Service

- Reviewer for: ICML, NeurIPS, ICLR, UAI, AISTATS.
- Past reviewer for: NeuroImage, ICASSP, EUSIPCO.

## Education & Earlier Experience

**2016–2017** **Ecole Polytechnique**, Palaiseau, France – Master of Research in Data Science; machine-learning-oriented curriculum.

**2015–2016** **University Lille 1**, Villeneuve d'Ascq, France – Master of Research in Applied Mathematics; focus on probability and statistics.

**2012–2016** **Ecole Centrale de Lille**, Villeneuve d'Ascq, France – Master of Engineering, Diplôme d'ingénieur; mathematics and computer science.

**2017** **Research Assistant Intern**, PARIETAL / NeuroSpin, CEA-Saclay – optimized MRI reconstruction pipelines using compressed sensing.

**2016** **Research Engineer Intern**, MODAL, INRIA Lille – developed an R package for clustering categorical functions with an efficient `C++` back-end.

**2015** **Software Developer Intern**, DataRobot, Boston – implemented multi-threading support for decision-tree induction in an AutoML system.