Yentl COLLIN

LinkedIn: linkedin.com/in/yentl-collin Email: yentl.collin@eleves.enpc.fr Mobile: +33 7 83 48 05 83

Education

ENS Paris-Saclay – Master's Degree MVA (Mathematics, Vision, Learning) Paris-Saclay, France
Main courses: Advanced Learning for Text/Graph Data; Optimal Transport; ML for Time Series; Probabilistic Graphical Models; Computational Statistics; RL; XAI; Representation Learning; LLM large-scale training. 2025–Present

Engineering Master's Degree – Ecole Nationale des Ponts et Chaussées Paris, France
Department of Mathematics, Probability and Computer Science. Main courses: Machine Learning; Computer Vision; Advanced Programming; Stochastic Processes; Statistics. 2022–Present

Lycée Pierre de Fermat – Preparatory Classes for Engineering Schools (MPSI/PSI) Toulouse, France
Two-year intensive program in Mathematics and Physics; ranked top 3% nationwide. 2020–2022

Publications

IFM Internship – Distill Locally, Schedule Globally: Flow Maps for Few-Step Text-to-Speech
First-author paper – Submitted to ICASSP 2027 – Paper link

Harvard Internship – SSM-CGM: Interpretable State-Space Forecasting Model of Continuous Glucose Monitoring for Personalized Diabetes Management
Accepted first-author paper – NeurIPS 2025 – TS4H Workshop – Paper link

Technical reports and other research papers from the IFM–MBZUAI internship will be released soon.

Experiences

AI Scientist Intern – Institute of Foundation Models (MBZUAI) – Speech / LLM Team Paris, France
Research on large-scale generative foundation models, with exclusive access to up to 300 nodes of 8 GPUs. 2026–Present

FLOW MAPS FOR TTS: Flow-map distillation of large-scale flow-matching TTS models for few-step inference.

DISCRETE DIFFUSION FOR ASR: Adapted a pretrained 7B autoregressive ASR backbone into a discrete-diffusion speech-to-text model, achieving 4× faster decoding at comparable accuracy.

INFERENCE-TIME SCALING FOR DRIFTING: Developing self-induced flow maps to enable inference-time scaling in drifting-based image generation.

LARGE-SCALE LLM TRAINING: Distillation of quadratic Transformers into subquadratic SSMs; agentic-data SFT and discrete-diffusion LLM training; RL for discrete-diffusion image models.

Contributed to training the Jais 3 (Link) language models and the Jais-ASR (Link) speech recognition model.

Visiting Student Researcher - Harvard University - Chirag J. Patel Lab - DBMI Boston, USA
Research on deep state-space models and attention-based architectures for time-series forecasting and anomaly detection, with a focus on explainability (using Mamba-2 and PyTorch Forecasting codebases). 2025–2026
Explored model-based RL for glucose management, leveraging our forecasting model as a learned simulator.
Attending courses: AI in Medicine II (Harvard, Marinka Zitnik); ML for Healthcare (MIT 6.7930).

ML Research Intern – Ida Paris, France
Conducted ML research on time-series learning for large-scale retail. Shipped end-to-end model pipelines. 2024–2025

Mathematics Tutor Paris, France
Taught mathematics to high school, undergraduate, and master's students. 2022–Present

Projects

LLM Engineering
From-scratch LLM: Built and trained a PyTorch decoder-only MoE Transformer from scratch, fine-tuned it with QLoRA, and applied GRPO on mathematical reasoning tasks. Implemented a hybrid Mamba/Transformer ASR model.

Paper Replications and Extensions
PyTorch reimplementations and extensions: SamFormer (improved optimizer/training stability), TabPFN (retrieval to improve long-context behavior), Mamba 2 (for speech recognition), Hyperspherical VAE (applied to EEG time-series), MOHAWK (Transformer-to-SSM distillation), Retrieval-Aware Distillation (Transformer–SSM hybrids), Sleeper Agents (email sending).

Hackathon World Model - Hack the World
Developed a JEPA-based world model predicting multiple action-conditioned future latent states in one forward pass.
Implemented self-speculative MPC rollouts with trajectory drafting and verification, achieving 2× faster inference than the AR.

Interests and Skills Summary

Music Experience: Diploma – Award – Regional Conservatoire of Toulouse Toulouse, France
Studied cello and music theory at a high level for 14 years. 2007–2021

Sport Experience: Soccer – Swimming – University-level soccer and swimming competition.

Languages: Native French, fluent English, notions of German, Hebrew, Italian, and Japanese.
Technical Skills: Python, PyTorch, C++, R, large-scale multi-node training.