Vincent PAULINE

Availability : Anytime – **Expected PhD Graduation :** July 2028
vincent.paulinef@gmail.com
google scholar - website

EDUCATION

MSc CentraleSupélec – Top 2 French Engineering School part of Paris-Saclay University (Top 15 Shanghai Ranking) 2020 - 2024
Master of Science & Engineering GPA : 4.11/4.33
Probability, Statistics and Machine Learning

MSc Computational Neuroscience and Neuroengineering Paris-Saclay University 2023 - 2024
Master thesis: "Towards a neurocomputational theory of inner speech"
Supervisors: Dr. Ladislas Nalborczyk – Prof. Dr. Thomas Hueber GPA : 4.33/4.33
Deep Learning - Neuroscience

Non-degree : Advanced Biology Course in Computational Neuroscience ENS Ulm PSL
Advanced Biomedical Course Gustave Roussy Cancer Research 2023 – 2024

Classe Préparatoire Lycée Lakanal (France) 2018-2020
Intensive two-year study course preparing for the competitive entrance examinations to the French 'Grandes Écoles' (the top French and highly-selective institutions) GPA : 4.33 /4.33 Valedictorian

EXPERTISE

Quantitative	Probability theory, Statistical inference, Deep learning, Optimization, Generative AI
Technical Skills	Python, R, Matlab, Docker, SQL, Java
Foreign Languages	English (Proficient), French (Native Speaker), Spanish (Basic)

PROFESSIONAL EXPERIENCE

Helmholtz AI - Pr. Dr. Stefan Bauer's lab Researcher Oct 24-July 25
(Munich, Germany) PhD student July 24-Now
PhD student – MCML Junior Fellow

- Working on the mathematical foundations of generative modeling
- Expanding diffusion and flow methods to general state spaces (text, sequences, multi-modal), with applications in sampling, guidance, fine-tuning, and diffusion large language models.
- First author of a 100-page general introduction to the field, with international collaborators Prof. Kirill Neklyudov (Anthropic, MILA, Canada) and Prof. Alexander Tong (AITHYRIA, Vienna): *"Foundations of Diffusion Models in General State Spaces: A Self-Contained Introduction."*
- Built the first unified codebase for training large diffusion language models: *"UNI-D²: A Unified Codebase for Discrete Diffusion Language Models."*

UNICOG - NeuroSpin Pr. Dr. Stanislas Dehaene's lab– 1 year Part time Sep 23-Apr 24
(Gif sur Yvette, France) Full Time Apr-Sept 24
Deep learning researcher

- Developed self-supervised generative speech models to replicate the mechanisms of both overt and inner speech
- Enhanced proficiency in designing and implementing advanced deep learning models, with a focus on applications in computational neuroscience

OWKIN	**Owkin – Dr. Mathieu Galtier's team – 10 months** **(Paris, France)** Data platform manager • Contributed to the development of Abstra, an AI-driven data platform, that facilitates both local and remote work on multimodal datasets using Federated Learning • Led the design and implementation of a flexible obfuscation toolchain for AI/ML models, coordinating with ML engineers from Romania, India, and Chile	Full time Feb-July 23 Part time Aug-Dec 23
BIOSERENITY Smart Healthcare Solutions	**BioSerenity (Maryland, United States) - 6 months** Industrialisation of AI-assisted diagnostic medical devices (EEG-PSG) • Set up the design & industrialization strategy in coordination with France, the US and China • Mapped and implemented the new automatic logistic process • Managed the production and quality testing of a new device entering the U.S. market, successfully facilitating Letter to File and 510(k) clearance	Full time Aug-Feb 23
	STMicroelectronics (Paris, France) – 10 months Development lead, IoT-oriented blockchain • Developed an IoT oriented blockchain presented to STM engineering lead • Pioneered an innovative solution to secure communication and data integrity	Part time Sep 21-June 22

ACADEMIC PROJECTS

Private Teacher	• Since 2020, taught six hours per week of private lessons in Algebra and Mathematical Analysis to undergraduate and graduate students
Deep Learning Kaggle Challenges	• Participated in a series of five Kaggle competitions at CentraleSupélec, including tasks in image classification, video classification, text classification, image generation and function fitting • Achieved an excellent second-place finish overall
Advanced course on high-tech tool design	• Prototyped an insertion device for a synchrotron beamline to expose it to researchers • Handled the specifications to allow angiography of frog ovary cells
Finalist European BEST Eng. Competition	• Shaped a 3-year business model for Saint-Gobain FRANCE • Designed an innovative drone for Venus exploration • Secured second place in the European case study competition

PERSONAL ACHIEVEMENTS

MCML Junior Fellow	• Awarded the MCML PhD Fellowship. One of 6 candidates selected from over 1,000 applications
Swimming marathon	• 2021, won a swimming marathon, 24 hours charity event for ATP fourth world • Achieved a new championship record
French merit-based scholarship	• National scholarship awarded for academic excellence to students from less privileged backgrounds, supporting my university studies.
French swimming team captain	• Led the French high school swimming team, fostering team cohesion • Achieved second place at the European Championships

SELECTED WORK

Foundations of Diffusion Models in General State Spaces: A Self-Contained Introduction. - paper
Vincent Pauline, Tobias Höppe, Kirill Neklyudov, Alexander Tong, Stefan Bauer, Andrea Dittadi

Self-contained tutorial unifying continuous diffusion (SDEs) and discrete diffusion (CTMCs) through infinitesimal generators, with common derivations of forward dynamics, time reversal, and variational training objectives.

One for All, All for One: Coordinated Multi-Agent Diffusion Steering via Stochastic Optimal Control. - paper
Riccardo Barbano, Vincent Pauline*, Runchang Li*, George Webber, Alexander Denker, Željko Kereta, Stefan Bauer, Francisco Vargas, Esmeralda S. Whitammer*

A stochastic optimal control framework that learns to coordinate frozen pretrained diffusion models without joint demonstrations, with applications to robotics, human motion, and inverse problems. Achieved 90% joint success versus 75% for PCD++, with >250× faster sampling on four-person motion coordination.

Amortised Inference through One-Step Implicit Sampling. - paper
Vincent Pauline, Kirill Tamogashev*, Arran Carter, Sanghyeok Choi, Stefan Bauer, Esmeralda S. Whitammer*

Trains one-step neural samplers from unnormalised densities through MCMC distillation, without target datasets. Connects idealised training dynamics to Wasserstein gradient flows. Achieves up to 1,000× faster sampling than tested baselines on ManyWell at comparable W_2^2 error, with no energy evaluations at inference.

UNI-D²: A Unified Codebase for Discrete Diffusion Language Models. - code
Kalyan Varma Nadimpalli, Vincent Pauline*, Ferdinand Kapl, Amir Mohammad Karimi-Mamaghan, Alexander Tong, Andrea Dittadi, Stefan Bauer*

An open-source PyTorch framework unifying prominent discrete diffusion language modelling approaches. Combines modular training, sampling, and evaluation with distributed mixed-precision training and FlashAttention integration, enabling reproducible benchmarking and rapid experimentation.

DDNO: Discrete Diffusion Noise Optimization. - paper
Luca Eyring, Vincent Pauline*, Stefan Bauer, Alexey Dosovitskiy, Zeynep Akata*

Steers frozen discrete diffusion models at test time by optimising initial noise through a differentiable surrogate combining straight-through gradients and soft mixing. Outperforms guidance and Best-of-N baselines on controllable text generation and compositional image synthesis.

Coming soon: Pushing the frontier of MoE diffusion language models by solving the continuous batching dilemma and the problem of expert explosion.
