Giorgio Giannone

ggiorgio@mit.edu — gigi@dtu.dk

linkedin/giorgio-c-giannone github/georgosgeorgos

I am a Principal Research Scientist on the AI Innovation Team at **Red Hat** in Boston and hold an appointment as a Research Affiliate at **MIT MechE**.

I work broadly in Generative AI and Probabilistic Methods, with a focus on **Inference-Time Scaling**, **Test-Time Adaptation**, **Vision-Language Alignment**, and **Few-Shot Generation**.

Experience

Principal Research Scientist, Red Hat, an IBM Company — Boston, Massachusetts, USA
June 2025 - Present

- AI Innovation Team
  - Probabilistic Inference for Vision and Language Models
  - Inference-Time Scaling and Reasoning for LLMs
  - Context Optimization for AgentOps

Research Affiliate, Massachusetts Institute of Technology — Cambridge, Massachusetts, USA
2025 - Present

- DeCoDE Lab, Department of Mechanical Engineering
  - Inference-Time Scaling for Constrained Generative Design
  - Iterative Self-Training for CAD Program Synthesis

Applied Scientist, Amazon — Seattle, Washington, USA
2024 - 2025

- Home Innovation and GenAI Team
  - Grounded Vision-Language Models
  - Subject-Driven Generative Models
  - Evaluation for Text-to-Image Models

Visiting Researcher, UCL Centre for Artificial Intelligence — London, UK
Jan 2024 - March 2024

- Host: David Barber
  - Multi-Resolution Convolutional Models for Long Sequences
  - Bayesian Inference for Language Models

Researcher (PhD Intern), Microsoft Research — Cambridge, Massachusetts, USA
Jun 2023 - Sept 2023

- ML and Statistics Group. Hosts: David Alvarez Melis, Nicolo Fusi
  - Dynamic Vocabulary Augmentation for LLMs

Research Collaborator, MIT-IBM AI Lab — Cambridge, Massachusetts, USA
Jan 2023 - June 2023

- Model Alignment Team. Host: Akash Srivastava
  - Generative Models for Systems with Constraints
  - Aligning Language Models with Negative Data

Research Scientist (PhD Intern), IBM Research — Zurich, Switzerland
Jun 2022 - Nov 2022

- Accelerated Discovery Team. Hosts: Matteo Manica, Teodoro Laino
  - Open-source library GT4SD for conditional generative models
  - Multitask Language Models for Text and Chemistry

Applied Scientist (PhD Intern), Amazon Science — Cambridge & London, UK
Jul 2021 - Oct 2021

- Alexa Team. Hosts: Yunlong Jiao, Emine Yilmaz
  - Domain Agnostic Subpopulation Generalisation

Research Engineer, NNAISENSE Lugano, Switzerland
Jan 2019 - Jan 2020

- Deep Learning Team. Managers: Christian Osendorfer, Jonathan Masci
  - Structured Latent Variable Models

Machine Learning Engineer, Pi Campus Rome, Italy
Oct 2018 - Dec 2018

- NLP for large scale data-driven early stage investing

Research Intern, Naver Labs Europe Grenoble, France
Feb 2018 - Aug 2018

- Computer Vision Team. Host: Boris Chidlovskii
  - Deep Learning for Semantic Scene Understanding and Mobile Robotics

Co-Founder, SecretAIry (formerly GAiA) Rome, Italy
July 2017 - Jan 2019

- Chatbots to enhance Workplace Communication
  - Selected among 100+ startups to join the EnLabs Incubator

Education

PhD, Generative Machine Learning Technical University of Denmark, Lyngby, Denmark
June 2020 - Dec 2023

- Few-Shot Generative Models
- Hierarchical Variational Inference
- Multitask Language Models for Conditional Molecule Generation
- Diffusion Models for Generative Engineering Design and Topology Optimization
- Thesis: Learning Generative Models with Limited Data
  - Supervisor: Ole Winther; Co-supervisor: Søren Hauberg

Visiting PhD Student, MIT School of Engineering Cambridge, Massachusetts, USA
Jan 2023 - Sept 2023

- Constrained Diffusion Models for Engineering Design (NeurIPS & Patent)
- Improving Generative Constraint Satisfaction using Invalid Designs (TMLR)
- Evaluating Vision-Language Models for Engineering Tasks (Journal)
- Research on LLM Agents for CAD design. Co-developer of `text2cad`.
  - Host: Faez Ahmed, DeCoDE Lab

Master's Degree, Data Science Sapienza University, Rome, Italy
Sept 2016 - Nov 2018

- Excellence Path & Summa Cum Laude
- Thesis: Multimodal Learning for Scene Understanding
  - Supervisor: Aris Anagnostopoulos; External Supervisor: Boris Chidlovskii

Visiting Graduate Student, NYU Tandon School of Engineering NYC, New York, USA
Sept 2017 - Jan 2018

- Visualization and Data Analytics Research Center. Host: Enrico Bertini
  - Built an interactive entity retrieval tool to investigate 10M documents

Master's Degree, Mechanical Engineering Sapienza University, Rome, Italy
Sept 2014 - Jan 2017

- Summa Cum Laude
- Thesis: Bubble Dynamics in Turbulent Shear Flows
  - Supervisor: Carlo Massimo Casciola; Co-supervisor: Paolo Gualtieri

Bachelor's Degree, Mechanical Engineering Sapienza University, Rome, Italy
Sept 2009 - May 2014

- Thesis: Rapid Prototyping of Metallic Manufacturing

Publications & Patents

Mitigating Premature Exploitation in Particle-based Monte Carlo for ITS under-review
Giannone, Xu, Nayak, Awhad, Sudalairaj, Xu, Srivastava *2025*

Generative optimization models for machine learning US Patent (MIT & IBM)
Giannone, Srivastava, Ahmed *2025*

Feedback-Driven Vision-Language Alignment under-review
Giannone, Li, Feng, Perevodchikov, Chen, Martinez *2025*

Be More Specific: Evaluating Object-centric Realism in Synthetic Images CVPR
Liang, Corneanu, Feng, Giannone, Martinez *2025*

Evaluating Vision-Language Models for Engineering Design Springer Artificial Intelligence Review
Picard, Edwards, Doris, Mann, Giannone, Alam, Ahmed *2025*

NITO: Neural Implicit Fields for Resolution-free Topology Optimization TMLR
Nobari, Regenwetter, Giannone, Ahmed *2025*

Reparameterized Multi-Resolution Convolutions for Long Sequence Modelling NeurIPS
Cunningham, Giannone, Zhang, Deisenroth *2024*

Constraining Generative Models for Engineering Design with Negative Data TMLR
Regenwetter, Giannone, Srivastava, Gutfreund, Ahmed *2024*

Aligning Optimization Trajectories with Diffusion Models NeurIPS
Giannone, Srivastava, Winther, Ahmed *2023*

Diffusing the Optimal Topology: A Generative Optimization Perspective IDETC23
Giannone, Ahmed *2023*

Unifying Molecular and Textual Representations via Multi-task LM ICML
Christofidellis*, Giannone*, Born, Winther, Laino, Manica *2023*

Accelerating Material Design with GT4SD Nature npj Computational Materials
GT4SD Team (Core Contributor) *2023*

Few-Shot Diffusion Models SBM@NeurIPS
Giannone, Nielsen, Winther *2022*

SCHA-VAE: Hierarchical Context Aggregation for Few-Shot Generation ICML
Giannone, Winther *2022*

Method and apparatus for semantic segmentation and depth completion US Patent (NAVER)
Chidlovskii, Giannone *2022*

JM1: Worst-group Generalization by Group Interpolation NeurIPS-W
Giannone, Havrylov, Massiah, Yilmaz, Jiao *2021*

Hierarchical Few-Shot Generative Models NeurIPS-W
Giannone, Winther *2021*

Transformation-aware Variational Autoencoders Technical Report
Giannone, Saremi, Masci, Osendorfer *2020*

Input-filtering NeuralODEs for spiking data NeurIPS-W
Giannone, Anoosheh, Quaglino, D'Oro, Masci, Gallieri *2020*

$\mathcal{T}$-VAE: No Representation without Transformation NeurIPS-W
Giannone, Masci, Osendorfer *2019*

Learning Common Representation from RGB and Depth Images CVPR-W
Giannone, Chidlovskii *2019*

Open-source

its-hub: A Python library for inference-time scaling 2025

- Inference-Time Scaling for Language Models.
- Focus on Mathematical Reasoning.
- Contributed Entropic Particle Filtering algorithms and new benchmark.

GT4SD: Generative Toolkit for Scientific Discovery 2022

- Library leveraging conditional generative models for accelerated discovery.

- Core Contributor.
- Work on Diffusion Models for images and 3D molecule conformation. The GFlowNet framework. Property Prediction module. Public Hub for model upload. Training Pipelines. Documentation. Tutorials. Testing. CI/CD. Server and Client API. Docker Images for CPU and GPU.

Grants & Awards

GPU Grant, LUMI-G, EuroHPC — Copenhagen, Denmark
PI, Efficient Pre-training of Large Generative Models for Constrained Design — *Nov 2023*

Grant, Otto Mønsted's Foundation — Copenhagen, Denmark
Grant Research Abroad — *Dec 2022*

Grant, Independent Research Fund Denmark — Lyngby, Denmark
DFF PhD Grant — *Jun 2020*

Grant, Perception as Generative Reasoning Workshop — NeurIPS 2019
Complimentary Conference Registration — *Oct 2019*

Grant, Pi School — Rome, Italy
Full Tuition for the School of AI (3% acceptance rate) — *Oct 2018*

Certificate of Award, Tsinghua University — Beijing, China
Prize for Outstanding Accomplishments, Deep Learning Summer School — *Aug 2018*

Certificate of Achievement, Naver Labs Europe — Grenoble, France
Prize for the Best Internship Performance — *Jul 2018*

1st Pick, Excellence Path, Master's Degree, Data Science — Rome, Italy
Admission based on the First year's Academic Performance — *Mar 2018*
Participation in the School for Advanced Studies

1st Place, Global AI Hackathon, Italian Edition — Rome, Italy
Our team built GAiA, an Enterprise Chatbot Assistant — *Jun 2017*
We won three prizes: Challenge Microsoft, People's Choice, Product Market Fit

Academic Service

Reviewer

Conference: ICML19, ICCV19, AAAI20, ICML21 (top 10%), AISTATS21, ICML22, NeurIPS22, CVPR23, NeurIPS23, ICML24, ICLR25, CVPR25, NeurIPS25, ICLR26

Journal: TPAMI, TMLR

Workshop: NeurIPS-IBW20, NeurIPS-MetaLearn21, ICML-DeployableGenAI23, ACL-LanguageMolecules24

Teaching

Teaching: Deep Learning (DTU 02456), Bayesian Machine Learning (DTU 02477), Advanced Machine Learning (DTU 02460)

Supervision: two special courses (9 months), two master's thesis (6+6 months), 18 final projects

Volunteering

PAISS18, NeurIPS18, ELLIS Unit Copenhagen, MLLS

Skills

Languages

- Python (proficient); R, Matlab (good knowledge); C, Java, JavaScript (basic knowledge)

Research

- Accelerate, HF Transformers, LaTeX, NLTK, OpenCV, PyTorch, SpaCy, TensorFlow

Software

- AWS, CVX, Docker/podman, FastAPI, Git, GitHub Actions, Gradio, Linux, MinIO, MongoDB, MySQL, Travis, vLLM

Miscellaneous

Summer/Winther Schools

- OxML22 , ProbAI21, M2L21, SMILES20, EEML20, RegML20, ETH School on PDEs, Tsinghua DL 2018, PAISS18

Talks

- Algorithmic Methods for Data Mining (Sapienza University), Bayesian Reading Group (DTU), MLLS Center (KU), UCL-NLP (UCL), Amazon Alexa (Cambridge), DeCoDE Lab (MIT)

Online Education

- Coursera: Machine Learning (Oct 2016), Deep Learning (Aug 2017).
- edX:

  Computer Science (Nov 2016), Artificial Intelligence (Apr 2017), CS50 (Jan 2021),
  Math for Quant Finance (Oct 2021), Causal Diagrams (Nov 2021), Science and Business of Biotech (Jun 2022).
- Udacity: Self-Driving Cars Nanodegree, 1st term (Dec 2017).

Associations/Communities

- Italian Association for Machine Learning (IAML)
- ContinualAI
- TribeAI