

Irina Espejo Morales

Research Scientist in AI for Science

📍 New York, NY ✉ irina.espejo@nyu.edu 🔗 irina-espejo 🌐 irinaespejo 📖 Google Scholar ✖ irinespejo

Education

PhD	New York University , Data Science • Advisor: Kyle Cranmer. • Scholarships: Fulbright Scholarship, DeepMind Fellowship, IRIS-HEP Fellowship.	
MSc	University of Oxford , Mathematical & Theoretical Physics	2017 - 2018
BSc	Universitat Autònoma de Barcelona , Mathematics	2012 - 2017
BSc	Universitat Autònoma de Barcelona , Physics	2012 - 2017

Experience

Research Scientist , Polymathic AI & New York University • Developing the NMR Agent for molecular-structure elucidation and mentoring summer interns on AI-for-science projects spanning chemistry, biology, and agentic systems.	NYC 05/26-now
Postdoctoral Researcher , Polymathic AI & New York University • Core contributor to MIMIC, a 1B-parameter generative multimodal foundation model for biomolecules, pretrained on 92 H200 GPUs; contributed to architecture, modality alignment, and RASP/context evaluations.	NYC 05/25-05/26
Postdoctoral Researcher, Accelerated Discovery & AI , IBM Research • Developed models and multimodal methods for chemistry, including encoded lookup-table modeling, automated data analysis, and laboratory activity recognition.	Zurich, CH 10/24-04/25
Research Scientist Intern, Accelerated Discovery & AI , IBM Research • Fine-tuned a 4-bit 7B model for chemistry tool use on approximately 300 examples, matching IBM's Granite 20B zero-shot performance on a laptop; also improved a laboratory video model by 7%.	Zurich, CH 11/23-09/24
Machine Learning Intern , BigHat Biosciences • Evaluated active learning, multi-objective Bayesian optimization, MCMC, and diffusion models for wet-lab-guided antibody design.	Bay Area 06/23-08/23
Research Assistant; Participant Scientist , New York University & ATLAS Experiment (CERN) • Built active-learning and multi-task Gaussian-process methods for new-particle searches, reducing confidence-interval data needs by 100x and integrating them with ATLAS simulation workflows.	NYC 2018 - '23
Student Visitor , Brookhaven National Laboratory • Automated ML-based parameter-inference workflows for high-energy physics, reducing runtime from days to five hours in collaboration with NERSC.	2020 - '21
Summer Research Intern , ATLAS Experiment, CERN • Applied random forests, boosted trees, and MLPs to highly imbalanced particle-physics data, improving sensitivity to new signals for the high-luminosity LHC.	Geneva, CH 06/17-08/17

Awards

Fulbright Scholarship

Awarded in 2017

Spanish Fulbright Commission

- Awarded for academic years 2018-2020. Only 25 scholarships per year are awarded across all disciplines. Approximate total value: \$200K (two years of NYU tuition and stipend).

DeepMind Fellowship

Awarded in 2018

New York University

- Awarded for academic year 2018-2019. Approximate total value: \$100K (one year of NYU tuition and stipend).

IRIS-HEP Fellowship

Awarded in 2021

- Awarded for academic year 2021-2022. Approximate total value: \$100K (one year of NYU tuition and stipend).

Publications

NMR Elucidation as an Agentic Search Problem, Not a Modeling Problem [paper](#)

[Irina Espejo Morales](#), Damon J. Hinz, Marvin Alberts, Geraud Krawezik, Haewon Jeong, Shirley Ho
ICML 2026 Workshop AI4Science (Spotlight Talk), 2026

MIMIC: A Generative Multimodal Foundation Model for Biomolecules [paper](#)

Siavash Golkar, Jake Kovalic†, [Irina Espejo Morales](#)†, Samuel Sledzieski†, Minhuan Li†, Ksenia Sokolova, Geraud Krawezik, Alberto Bietti, Claudia Skok Gibbs, Roman Klypa, Shengwei Xiong, Francois Lanusse, Liam Parker, Kyunghyun Cho, Miles Cranmer, Tom Hehir, Michael McCabe, Lucas Meyer, Rudy Morel, Payel Mukhopadhyay, Mariel Pettee, Helen Qu, Jeff Shen, David Fouhey, Hadi Sotoudeh, Vikram Mulligan, Pilar Cossio, Sonya M. Hanson, Alisha N. Jones, Olga G. Troyanskaya, Shirley Ho
arXiv preprint, 2026

† Equal contribution

Semantic Search for 100M+ Galaxy Images Using AI-generated Captions [paper](#)

Nolan Koblishke, Liam Parker, Francois Lanusse, Jo Bovy, [Irina Espejo](#), Shirley Ho
The Astrophysical Journal, 2026

Efficient search for new physics using Active Learning in the ATLAS Experiment [paper](#)

[Irina Espejo](#), Patrick Rieck and the ATLAS Collaboration
Journal of Physics: Conference Series, 2026

Scaling MadMiner with a deployment on REANA [paper](#)

[Irina Espejo](#), Sinclert Pérez, Kenyi Hurtado, Lukas Heinrich, Kyle Cranmer
Journal of Physics: Conference Series, 2026

Unified lookup tables: training foundation models on encoded data [paper](#)

Nikita Janakarajan†, [Irina Espejo Morales](#)†, Marvin Alberts, Andrea Giovannini, Matteo Manica, Antonio Foncubierta-Rodríguez
Machine Learning: Science and Technology, 2025

† Equal contribution

Making Sense of Data in the Wild: Data Analysis Automation at Scale [paper](#)

Mara Graziani, Malina Molnar, [Irina Espejo Morales](#), Joris Cadow-Gossweiler, Teodoro Laino
arXiv preprint, 2025

Activity recognition in scientific experimentation using multimodal visual encoding [paper](#)

Gianmarco Gabrieli, [Irina Espejo Morales](#), Dimitrios Christofidellis, Mara Graziani, Andrea Giovannini, Federico Zipoli, Amol Thakkar, Antonio Foncubierta, Matteo Manica, Patrick W. Ruch
Digital Discovery, 2025

Interleaving Text and Number Embeddings to Solve Mathematics Problems [paper](#)

Marvin Alberts, Gianmarco Gabrieli, [Irina Espejo Morales](#)

arXiv preprint, 2024

MadMiner: Machine Learning-Based Inference for Particle Physics [paper](#)

Johann Brehmer, Felix Kling, [Irina Espejo](#), Kyle Cranmer

Computing and Software for Big Science, 2020

Constraining effective field theories with machine learning [paper](#)

Johann Brehmer, Kyle Cranmer, [Irina Espejo](#), Alexander Held, Felix Kling, Gilles Louppe, Juan Pavez

EPJ Web of Conferences, 2020

Skills

Software Engineering

Highlighted repos and contributions:

- madminer-workflow: reduced runtime of complex workflows integrating ML and physics simulators from days to 5 hours.
- Contributed to maintaining ONMT training framework for rxn-chemistry forward and retro models.

Technologies

- Python, PyTorch, GPyTorch, PyTorch Lightning, BoTorch, Hugging Face, DeepSpeed.
- Git version control, CI/CD, multi-device distributed training, open source software, reusability, reproducibility, Docker, Singularity, Unix/Linux, MLflow, cloud technologies, bash, Slurm, GPUs, MLOps.

Languages

English: Full Business Proficiency

Spanish: Native

Catalan: Native