Gaia Grosso

Postdoctoral Associate, Laboratory for Nuclear Science
Massachusetts Institute of Technology

gaiag795@mit.edu · ggrosso@seas.harvard.edu
gaiagrosso.github.io
Google Scholar · INSPIRE · ORCID · GitHub

Research interests. Statistical methods for detecting when data deviate from a reference model in high-dimensional settings: two-sample and goodness-of-fit testing with calibrated error control, interpretable kernel and physics-inspired models, and hypothesis testing in learned representation spaces. Applications to scientific discovery in particle physics and to auditing, monitoring and interpreting large AI models.

Academic Appointments

Sep 2026 – present	**Postdoctoral Associate**, Laboratory for Nuclear Science, Massachusetts Institute of Technology
Sep 2023 – Aug 2026	**IAIFI Postdoctoral Fellow**, NSF AI Institute for Artificial Intelligence and Fundamental Interactions *Laboratory for Nuclear Science, MIT · School of Engineering and Applied Sciences, Harvard University* Mentors: Prof. Philip Harris (MIT), Prof. Demba Ba (Harvard).
Nov 2019 – Jan 2023	**CERN Doctoral Student**, CMS Experiment, CERN, Geneva, Switzerland ML-based signal-agnostic searches with systematic uncertainties; graph neural networks for fast jet reconstruction on GPUs; monitoring and control software for the CMS 40 MHz Level-1 scouting system. Supervisor: Dr. Maurizio Pierini.
Jun – Sep 2018	**CERN Summer Student**, CERN, Geneva, Switzerland Machine learning techniques to search for new physics (report). Supervisors: Dr. Maurizio Pierini, Dr. Olmo Cerri.

Education

2019 – 2023	**Ph.D. in Physics**, University of Padova, Italy (research carried out at CERN) Thesis: *Searching for unexpected new physics at the LHC with machine learning*. Developed the New Physics Learning Machine (NPLM) and deployed it on CMS data and detector monitoring; results in [J2–J5]. Advisors: Prof. Marco Zanetti, Prof. Andrea Wulzer.
2017 – 2019	**M.Sc. in Physics**, University of Padova, Italy · 110/110 *cum laude* Thesis: *Unbiased ML-based strategy for new physics searches at the LHC*; results in [J1]. Advisor: Prof. Marco Zanetti.
2014 – 2017	**B.Sc. in Physics**, University of Padova, Italy Thesis: *Deep learning techniques to search for new physics at the LHC*. Advisor: Prof. Marco Zanetti.

Fellowships, Funding and Awards

2025	**Breakthrough Prize in Fundamental Physics**, awarded to the ATLAS, CMS, ALICE and LHCb Collaborations; co-recipient as a member of CMS.
2023 – 2026	**IAIFI Fellowship**, NSF AI Institute for Artificial Intelligence and Fundamental Interactions. Competitively selected independent postdoctoral fellowship.
2019 – 2023	**CERN Doctoral Student Programme**. Competitively selected doctoral fellowship at CERN.
2018	**CERN Summer Student Programme**. Competitively selected.

Publications

Bold: my name. † Student I mentored. * Equal contribution. $^{\alpha}$ Authors in alphabetical order.

Peer-reviewed conference papers and proceedings (2)

[C2] S. Bright-Thonney, C. Reissel, **G. Grosso**, N. Woodward†, E. Govorkova, A. Novak, S.E. Park, E. Moreno, P. Harris. AutoSciDACT: Automated scientific discovery through contrastive embedding and hypothesis testing. *Advances in Neural Information Processing Systems* **38** (NeurIPS 2025), 88575–88614.

[C1] **G. Grosso**, N. Lai, M. Migliorini, J. Pazzini, A. Triossi, M. Zanetti, A. Zucchetta$^{\alpha}$. Triggerless data acquisition pipeline for machine learning based statistical anomaly detection. *EPJ Web Conf.* **295**, 02033 (2024). Proceedings of CHEP 2023.

Preprints and papers under review (6)

[P6] G. Badaracco†, S. Benevedes, C. Reissel, P. Harris, **G. Grosso**. An end-to-end framework for uncertainty-aware validation of generative models. arXiv:2610.37398 (2026).

[P5] S.X. Yu†, **G. Grosso**. Auditing fine-tuned models against benign variation. Under review (AISTATS 2027).

[P4] C. McNamee†, H. Orgad*, **G. Grosso***, N. Saphra, D. Ba. To be or not to be: mutually exclusive features in language models. Under review (ICLR 2027).

[P3] S. Tong†, P. Harris, **G. Grosso**. Covariant contrastive learning for uncertainty-aware anomaly detection. arXiv:2609.35755 (2026).

[P2] **G. Grosso**, S.S.R. Hindupur†, T. Fel, S. Bright-Thonney, P. Harris, D. Ba. Sparse, self-organizing ensembles of local kernels detect rare statistical anomalies. arXiv:2511.03095 (2025). Under review at *PNAS*.

[P1] N.S. Woodward†, S.E. Park, **G. Grosso**, J. Krupa, P. Harris. Product manifold machine learning for physics. arXiv:2412.07033 (2024).

Peer-reviewed journal articles (10)

[J10] P. Cappelli†, **G. Grosso**, M. Letizia, H. Reyes-Gonzalez, M. Zanetti. Learning to validate generative models: a goodness-of-fit approach. *Mach. Learn.: Sci. Technol.* **7**, 045011 (2026).

[J9] K. Metzger†, M. Sodini, L. Xu†, T. Aarrestad, E. Govorkova, **G. Grosso**, P. Harris. Anomaly-preserving contrastive neural embeddings for end-to-end model-independent searches at the LHC. *Phys. Rev. D* **112**, 072011 (2025).

[J8] **G. Grosso**, D. Sengupta, T. Golling, P. Harris. Robust resonant anomaly detection with NPLM. *Eur. Phys. J. C* **85**, 1074 (2025).

[J7] **G. Grosso***, M. Letizia*. Multiple testing for signal-agnostic searches for new physics with machine learning. *Eur. Phys. J. C* **85**, 4 (2025).

[J6] **G. Grosso**. Anomaly-aware summary statistic from data batches. *J. High Energy Phys.* **12** (2024) 093. *Single author.*

[J5] **G. Grosso**, M. Letizia, A. Wulzer, M. Pierini. Goodness of fit by Neyman–Pearson testing. *SciPost Phys.* **16**, 123 (2024).

[J4] **G. Grosso**, N. Lai, M. Letizia, J. Pazzini, M. Rando, L. Rosasco, A. Wulzer, M. Zanetti$^{\alpha}$. Fast kernel methods for data quality monitoring as a goodness-of-fit test. *Mach. Learn.: Sci. Technol.* **4**, 035029 (2023).

[J3] M. Letizia, G. Losapio, M. Rando, **G. Grosso**, A. Wulzer, M. Pierini, M. Zanetti, L. Rosasco. Learning new physics efficiently with nonparametric methods. *Eur. Phys. J. C* **82**, 879 (2022).

[J2] R.T. D'Agnolo, **G. Grosso**, M. Pierini, A. Wulzer, M. Zanetti$^{\alpha}$. Learning new physics from an imperfect machine. *Eur. Phys. J. C* **82**, 275 (2022).

[J1] R.T. D'Agnolo, **G. Grosso**, M. Pierini, A. Wulzer, M. Zanetti$^{\alpha}$. Learning multivariate new physics. *Eur. Phys. J. C* **81**, 89 (2021).

Workshop papers (*non-archival*) (3)

[W3] **G. Grosso**, A. Marin-Llobet, D. Ba. Adaptive and interpretable two-sample test as a batch-level tool for AI safety. ICML 2026 Workshop on Mechanistic Interpretability, Seoul.

[W2] G. Badaracco†, S. Benevedes, **G. Grosso**, C. Reissel, T. Aarrestad, P. Harris. An end-to-end pipeline for uncertainty-aware validation of generative AI. NeurIPS 2025 Workshop on Machine Learning and the Physical

Sciences, San Diego.

[W1] **G. Grosso**, E. Govorkova, P. Harris, E. Moreno, R. Raikman. Learning dictionaries of new physics with sparse local kernels. NeurIPS 2024 Workshop on Machine Learning and the Physical Sciences, Vancouver.

Reviews, perspectives and white papers (4)

[R4] **G. Grosso**, S. Rodriguez-Benitez†, L. Lyons. The look-elsewhere effect in searches for new physics: from classical methods to machine learning. Review, in preparation.

[R3] **G. Grosso**, R. Winterhalder, L. Brenner, L. Lyons, T. Plehn. VERAIPHY – Validation & Evaluation for Robust AI in PHYsics. Opening article of the PHYSTAT VERAIPHY series. arXiv:2608.17724 (2026).

[R2] **G. Grosso**, V. Mikuni, L. Heinrich. Are we ready for AI-driven discovery? AI verification before the next fundamental physics breakthrough. Review, VERAIPHY series. arXiv:2607.10039 (2026). To be submitted to *SciPost Physics*.

[R1] **G. Grosso**, P. Harris, S. Mishra-Sharma, P. Shanahan. A virtuous cycle: generative AI and discovery in the physical sciences. *An MIT Exploration of Generative AI* (2024).

Collaboration papers

Member of the CMS Collaboration since 2018; on the author list since 2021 and co-author of 465 CMS Collaboration papers (full list on INSPIRE-HEP).

Patents

2025	**G. Grosso**, D.E. Ba, S.S. Hindupur Ravindra†. Methods for explainable novelty discovery using artificial intelligence. U.S. Provisional Patent Application No. 63/909,442, filed October 31, 2025. Covers the SparKer method [P2].

Software

NPLM	New Physics Learning Machine: ML implementation of the likelihood-ratio test for signal-agnostic anomaly detection; calibrated p-values and localization of the most discrepant regions [J1–J5, J7, J8]. GitHub
SparKer	Sparse, self-organizing ensembles of local kernels for rare-anomaly detection in learned representations, with geometrically interpretable detection experts [P2, W3]. Code covered by a U.S. provisional patent application (see Patents).
Tutorials	Hands-on notebooks on diffusion models, VAEs and generative-model validation, written for the IAIFI Summer School 2024. GitHub

Invited Talks, Seminars and Panels

Sep 2026	*Discovery through the lens of AI: scaling anomaly detection with learned representations.* LPC Topic of the Week, LHC Physics Center, Fermilab, Batavia, IL. [event] [slides]
Apr 2026	*Detecting statistical anomalies with sparse, self-organizing ensembles of local kernels.* Statistical Physics for Ambitious Interpretability, Cambridge, MA. [event]
Oct 2025	*From high dimensions to statistical discovery: anomaly detection at scale in the era of latent spaces.* STAMPS–NSBI Workshop, Carnegie Mellon University, Pittsburgh, PA. [video]
Mar 2025	Panelist, round table on goodness-of-fit tests via machine learning. PHYSTAT Informal Review (remote). [event]
Sep 2024	*Statistical tests for anomaly detection at the LHC.* PHYSTAT – Statistics meets ML in Particle Physics & Astrophysics, Imperial College London, UK. [programme]
Nov 2023	*Uncertainty-aware ML models.* From Physics to Medicine: XAI Workshop. [event]
Sep 2022	Seminar: *New Physics Learning Machine (NPLM) at work.* Galileo Galilei Institute, Florence, Italy. [event]

Contributed Talks

Sep 2025	*Validation, monitoring and discovery: statistical tests for anomalies in the era of latent spaces.* FirstPrinciple Talks Program (remote).

Sep 2025	*Anomaly detection in the era of latent spaces.* Building Big or Building Smart, MIAPbP, TU Munich, Germany. [slides]
Jun 2025	*From high dimensions to statistical discovery: a contrastive learning approach to anomaly detection.* Anomaly Detection for HEP Workshop, Columbia University, Irvington, NY. [slides]
Aug 2024	*How good is your model? Goodness of fit by Neyman–Pearson testing.* IAIFI Summer Workshop, MIT, Cambridge, MA.
May 2024	*Boosting statistical anomaly detection via multiple testing with NPLM.* ACAT 2024, Stony Brook University, NY. [event]
Jun 2023	*Goodness of fit by Neyman–Pearson testing.* PHYSTAT-2sample (remote). [event]
May 2023	*Systematics in ML model-independent searches.* CHEP 2023, Norfolk, VA. [event]
Apr 2023	*Unbiased detection of data departures from expectations with machine learning.* Systematic Effects and Nuisance Parameters in Particle Physics Data Analyses, BIRS, Banff, Canada. [event]
Sep 2022	*Learning new physics from a machine.* Machine Learning at GGI, Florence, Italy. [event]
Aug 2022	*New Physics Learning Machine (NPLM) at work.* Hammers & Nails, Weizmann Institute of Science, Rehovot, Israel. [event]
May 2022	*Learning new physics aware of systematic uncertainties.* 5th Inter-experiment Machine Learning Workshop, CERN, Geneva. [event]
Apr 2022	*An imperfect machine to search for new physics: dealing with uncertainties in a machine-learning based signal extraction.* Learning to Discover, Institut Pascal, Paris-Saclay, France. [event]
Apr 2022	*Model-independent search for new physics at the LHC.* Spring Workshop on Physics of Data, Venice, Italy. [event]
Dec 2021	*An imperfect machine to search for new physics: a way to include systematic uncertainties.* ACAT 2021 (remote). [event]

Posters

Jul 2026	*Adaptive and interpretable two-sample test as a batch-level tool for AI safety* [W3]. ICML 2026 Mechanistic Interpretability Workshop, Seoul, South Korea.
Dec 2024	*Learning dictionaries of new physics with sparse local kernels* [W1]. NeurIPS 2024 ML4PS Workshop, Vancouver, Canada.
Mar 2024	*Building sparse kernel methods via dictionary learning: expressive, regularized and interpretable models for statistical anomaly detection.* EuCAIFCon 2024, Amsterdam, Netherlands. [event]
Dec 2022	*How good is the Standard Model? Machine learning multivariate goodness-of-fit tests.* NeurIPS 2022 ML4PS Workshop, New Orleans, LA. [paper]
Dec 2021	*An imperfect machine to search for new physics: systematic uncertainties in a machine-learning based signal extraction.* NeurIPS 2021 ML4PS Workshop (virtual).
	Co-authored posters presented by mentees and collaborators: Conference on Physics and AI (PAI 2026, Stanford) [P3]; NeurIPS 2025 main conference [C2] and ML4PS Workshop [W2]; EuCAIFCon 2025 (Cagliari); NeurIPS 2024 ML4PS Workshop [P1]; NeurIPS 2022 ML4PS Workshop.

Teaching

University courses

Mar 2026	Guest lecture, *Same or different? Classical and ML methods for two-sample testing in science.* CDS DS 595 AI Methods for Science, Boston University. [slides]
Apr 2025	Guest lecture, *Statistical tests for anomaly detection and scientific discovery.* 8.16/8.316 Data Science in Physics, MIT.
2022 – 2023	Teaching assistant, Laboratory of Computational Physics (Mod. A), M.Sc. program in Physics of Data, University of Padova.

Lectures and tutorials at schools

Feb 2027	Lecturer (upcoming), PHYSTAT School of Statistics, Damp, Germany (Feb 22 – Mar 1, 2027).

Nov 2025	Lecturer, *Statistical and ML tools for anomaly detection in fundamental physics.* First PHYSTAT School of Statistics: Statistics in the Era of Machine Learning, Nikhef, Netherlands.
Aug 2025	Lecturer, *Physics-guided optimizations: inform and inspire AI models with physics principles.* IAIFI Summer School, Harvard University. [slides]
Aug 2024	Tutorial developer and facilitator, *Diffusion models for physics*: designed notebooks on diffusion models, VAEs and validation; led the hands-on session. IAIFI Summer School, MIT. [notebooks]

Mentoring

14 students mentored since 2023 (4 undergraduate, 10 graduate and Master's) at MIT, Harvard, Princeton, ETH Zurich, University of Padova, IFT Madrid and IFAE Barcelona; 11 papers and manuscripts co-authored with mentees (marked † in the publication list). Primary research advisor unless noted; MIT undergraduates were supervised through the Undergraduate Research Opportunities Program (UROP). Linked project titles lead to the resulting papers.

Period	Student	Project
Current		
May 2026 –	Shirley X. Yu, undergraduate, Princeton	Auditing fine-tuned language models via representation drift
Feb 2026 –	Francesco Montagno, graduate, IFAE Barcelona	Anomaly detection in the four-lepton final state with CMS Open Data; interpretable EFT searches with contrastive latent spaces
Sep 2025 –	Cameron McNamee, graduate, Harvard (co-advised)	Interpreting large models with co-activation graphs [P4]
May 2025 –	Sergio Rodriguez-Benitez, graduate, IFT Madrid	Review of the look-elsewhere effect from classical methods to machine learning [R4]; regularization in ML-based anomaly detection
Apr 2025 –	Shelley Tong, graduate, MIT	Continuous systematic uncertainties in contrastive embeddings for LHC data; anomaly detection in the four-lepton final state with CMS Open Data
Former		
Mar – Aug 2026	Ahmed Fadhlallah, undergraduate, MIT	Metrics for evaluating embedding spaces for anomaly detection in the four-lepton final state with CMS Open Data
Aug – Oct 2025	Sai Sumedh R. Hindupur, graduate, Harvard	Theory of SparKer learning dynamics
Nov 2024 – Sep 2025	Giada Badaracco, Master's, ETH Zurich (co-advised)	Uncertainty-aware validation of generative AI
Sep 2023 – Sep 2025	Pietro Cappelli, graduate, Univ. of Padova	Generative-model validation via ML goodness-of-fit tests
Oct 2023 – Apr 2025	Lana Xu, undergraduate, MIT	Contrastive representations for anomaly detection
Apr 2023 – Apr 2025	Nathaniel S. Woodward, undergraduate, MIT	Product-manifold machine learning for particle physics
2023 – 2024	Nicolò Lai, graduate, Univ. of Padova	Signal-agnostic NPLM search in the CMS dimuon final state
Oct 2023 – Sep 2024	Kyle S. Metzger, Master's, ETH Zurich (co-advised)	Representation learning for new physics searches in CMS
Jan – Jun 2024	Jessica Prendi, Master's, ETH Zurich (co-advised)	Model-agnostic searches in the CMS 40 MHz scouting system

Leadership and Service

Leadership roles

2026 – present	**Co-organizer, PHYSTAT**: member of the event-organizer core team and of the Advisory Committee International series of workshops, seminars and schools on statistical methods in particle physics, bringing together physicists and statisticians (founded 2000). [webpage]

2025 – present	**Co-organizer, VERAIPHY** (Validation & Evaluation for Robust AI in Physics), PHYSTAT-endorsed initiative Article series on statistical standards for ML in fundamental physics [R2, R3]; workshop (Feb 2026). [webpage]
2025 – 2026	**Track convener**, "Software environment and maintainability," CHEP 2026, Bangkok, Thailand (May 2026). Abstract selection, session chairing, proceedings review.
2023 – 2025	**Convener, ML Innovation (L3)**, CMS Machine Learning Group, CERN Organized seminars, workshops and hackathons on emerging ML methods for CMS. Hackathons (internal): ML for Reweighting (Oct 2023); Anomaly Detection for Monitoring, Validation and Discovery (Jul 2024). Public workshops: *Foundation Models for Physics* (Oct 2024) and the *Bites of Foundation Models for Physics* series (Nov 2024, Jun 2025).
2024 – 2025	**Co-chair, IAIFI Speaker Selection Committee**: seminar and colloquium series.

Event organization

Mar 2025	Co-organizer, IAIFI Research Panel on the 2024 Nobel Prize in Physics, MIT. [video]
Nov 2024	Co-organizer, IAIFI Fireside Chat on the 2024 Nobel Prize in Physics, MIT. [video]
Mar 2024	Organizing committee, IAIFI Symposium on the Impact of Generative AI in the Physical Sciences, MIT; co-authored the resulting white paper [R1].

Refereeing

Journals	*PRX Intelligence*, *Physical Review D*, *Nuclear Physics B*, *Journal of High Energy Physics*, *European Physical Journal C*.
Conferences	ICLR, AISTATS, NeurIPS Workshop on Machine Learning and the Physical Sciences.

Outreach

Nov 2024	Cambridge Science Festival: built an LLM-powered, retrieval-augmented interactive Marie Curie persona answering visitors' questions in historically grounded form.
Nov 2024	Museum of Science, Boston: public talk and Q&A on AI and particle physics.
Sep 2023	Public lecture, *How good is the Standard Model? Hunting anomalies in the LHC data and beyond.* Institute of Advanced Study, University of Geneva. [event]
2023	CERN Science Gateway: video recording on physics at the LHC for the permanent exhibition.
Feb 2023	Panelist, *Women in Science 2023* (Donne nella scienza): round table for the International Day of Women and Girls in Science, University of Padova (800th-anniversary program). [event]
Nov 2020	Veneto Night: livestreamed presentation on new physics searches for the general public and children, University of Padova. [video]
2018	Built a cloud chamber with a student team for particle-detection demonstrations to high-school students, University of Padova.

Languages

Italian (native), English (fluent), Catalan (intermediate), French (basic).