Iñigo Urteaga

https://iurteaga.github.io

Research scientist working on probabilistic machine learning for descriptive, predictive, and prescriptive tasks. I have specialized in computational statistics, generative modeling, dynamical systems, approximate inference, and sequential decision processes. My research focuses on methodological and applied aspects of Bayesian Theory.

Professional Appointments and Experience

09/2025 **Ramón y Cajal Fellow**, *Basque Center for Applied Mathematics (BCAM)*, Bilbao, Spain

01/2023 **Ikerbasque Research Fellow**, *Basque Center for Applied Mathematics (BCAM)*, Bilbao, Spain
- Tenure-track researcher at BCAM's Machine Learning group.
  - Probabilistic machine learning and generative models: methodology and applications
  - Statistical modeling of time-varying phenomena, collected via not-at-random measurements.
  - Reinforcement learning, control theory, and multi-armed bandits with applications.

10/2024 – 12/2024 **Visiting Scholar**, *Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard*, Cambridge, Massachusets, (USA)
- Fundamentals and applications of probabilistic machine learning and dynamical systems
- Learning dynamics from partial and noisy observations via data-driven and mechanistic models
- Uncertainty quantification of dynamical linear systems: a Bayesian approach over reduced parameterizations

04/2018 – 12/2022 **Associate Research Scientist**, *Columbia University*, New York City, NY (USA)
- Descriptive modeling: *Bayesian deep generative modeling for mobile-health data.*
  - Phenotyping algorithms and models for healthcare data.
- Predictive modeling: *Accurate and robust prediction with uncertain and sparse measurement data.*
  - Gaussian processes and deep-learning for reconstruction and forecasting of hormonal dynamics.
  - Personalized, deep and generative models for health events with self-tracked mobile-health data.
- Prescriptive modeling: *Sequential decision making in complex practical scenarios.*
  - Multi-armed contextual bandits: theory and real-life applications.

09/2016 – 04/2018 **Postdoctoral Research Scientist**, *Columbia University*, New York City, NY (USA)
- Unsupervised phenotyping of endometriosis via self-tracked, smartphone based mobile health data.
- Mechanistic modeling and machine learning for the female hormonal cycle.
- Multi-armed contextual bandits: Thompson sampling and variational inference.

07/2009 – 07/2011 **Researcher**, *Tecnalia-Telecom*, Zamudio (Spain)
- SAIL: Human mobility analysis and pattern extraction. Complex network analysis.
- TelMAX: Mobile communication system design, multimedia applications and heterogeneous networks.

01/2009 – 07/2009 **Telecommunication Engineer**, *Traintic*, Donostia (Spain)
- Infrastructure-to-vehicle (I2V), Vehicle-to-vehicle (V2V) and On-Board networking on railways.

08/2007 – 06/2008 **Research Assistant**, *Colorado School Of Mines*, Golden, Colorado (USA)
- Wireless sensor networking and distributed systems (supervised by Dr. Qi Han)

04/2005 – 06/2007 **Research Scholarship**, *NQaS research group within (EHU/UPV)*, Bilbao (Spain)
- Network quality and service (supervised by Alex Muñoz Mateos)

Education

2011–2016 **Ph.D. in Electrical Engineering**, *Stony Brook University (USA)*, *GPA*: 3.97/4.0
- Dissertation
  - *Sequential Monte Carlo methods for inference and prediction of latent time-series.*
- Advisor: Prof. Petar M. Djurić
- Research Topics
  - Particle filtering for time-series: ARMA, FARIMA, fractional Gaussian processes.
  - Bayesian theory: parameter estimation, Rao-Blackwellization and hierarchical models.
  - Bayesian model selection and averaging.
  - Non-parametric Bayesian methods: Gaussian processes for regression and prediction.
  - Robust signal processing: outlier and missing data.
- Selected Courses
  - Stochastic processes, Probabilistic graphical models, Detection and estimation theory, Pattern recognition, Digital signal processing, Machine learning.

2002–2008 **M.S. Telecommunication Engineering**, *UPV-EHU*, Bilbao (Spain), *Grade*: 7.9/10
- Master Thesis (*Grade*: 10/10) supervised by Dr. Qi Han at Colorado School Of Mines (USA)
  - *REDFLAG: A Run-timE, Distributed, Flexible, Lightweight, And Generic Fault Detection Service for Data-Driven Wireless Sensor Applications.*

2000–2002 **Scientific Baccalaureate**, *Axular Lizeoa*, Donostia (Spain), *Grade: Honours*

Research projects and grants

03/12/2025 – 03/11/2028 **US, ONR: Basic and Applied Scientific Research**, *Award # N00014-25-1-2251*, $482,548.00, "*A Novel Approach to Design of Bandits: Deterministic Analytic Approximations for Bandit Moments*" PI Chris H. Wiggins (Columbia University), Co-PI Iñigo Urteaga.

09/01/2024 – 08/31/2028 **Spain, Agencia Estatal de Investigación, Plan Nacional**, *PID2023-146759NA-I00*, €93,000, "*Bayesian theory for non-stationary and multi-scale sequential decision processes*" PI Iñigo Urteaga.

09/01/2024 – 12/15/2024 **Spain, Centro de Tecnología de Repsol & Math-In**, €10,000, "*Surrogates for data-driven modeling of parametric complex systems*" PI Iñigo Urteaga.

09/01/2023 – 08/31/2027 **US, National Science Foundation, IIS - SCH**, *Award ID 2306690*, $1,197,325, "*Human-Centered Reinforcement Learning for Personalized Coaching in Health*" PI Lena Mamykina (Columbia University), Co-PI Iñigo Urteaga.

12/15/2022 – 12/14/2025 **LaCaixa Foundation's Junior Leader Incoming**, *Awarded, LCF/BQ/PI22/11910028*, €300,000, "*Statistical machine learning for real-life time-varying phenomena, collected via not-at-random measurement processes*". PI Iñigo Urteaga

03/01/2021 – 01/28/2023 **eBay Research & University Partnership for Technology**, *Awarded*, $125,000, "*Online optimization of Transformer-based Natural Language models: a Bandit based approach*" PI Iñigo Urteaga (Columbia University).

02/01/2019 – 01/31/2023 **US, National Institute of Health**, *R01 LM013043: Awarded*, $1,620,000, "*PhendoPHL: A Data-Science Enabled Personal Health Library to Manage Endometriosis*" PI Noémie Elhadad; Co-I Iñigo Urteaga (Columbia University).

02/15/2014 – 01/31/2020 **US, National Science Foundation**, *IIS-1344668: Awarded*, $1,994,224, SCH:INT "*Large-Scale Probabilistic Phenotyping Applied to Patient Record Summarization*" PI Noémie Elhadad, Co-PI Chris H. Wiggins, Investigator Iñigo Urteaga (Columbia University).

Awards and Recognitions

2021 **2021 STAT Wunderkind**, *In recognition of my early-career scientific work*
Significant contributions on statistical modeling and machine learning for mobile health data.

2025 **TMLR Expert Reviewer**, Recognition highlighting my contribution and expertise as reviewer for TMLR, for exemplary work in evaluating TMLR submissions.

2023 **AISTATS 2023 Top Reviewer**, *Amongst the top 10% highest-scoring reviewers*

2020 **NeurIPS Top Reviewer**, *Amongst the top 10% highest-scoring reviewers*

2019 **NeurIPS Top Reviewer**, *Amongst the 400 highest-scoring reviewers*

2018 **NeurIPS Top Reviewer**, *Amongst the 30% highest-scoring reviewers*

2016 **Best Graduate Student**, *Electrical and Computer Engineering at Stony Brook University*
Armstrong Memorial Research Foundation

Spring 2016 **Provost Graduate Lecture Series Speaker**, Stony Brook University, USA
Lecture available online at https://youtu.be/67KfUVXlkl0

Fall 2015 **Distinguished Travel Award for Fall 2015**, Stony Brook Graduate School

2015 **Professional Development Awards Program**, *Stony Brook University*
New York State and Graduate Student Employees Union

2009-2011 **Torres Quevedo Research Fellowship**, *PTQ-09-02-01814*, Robotiker-Tecnalia
Ministerio de Ciencia e Innovación, España

2007 **Global Education for European Engineers and Entrepreneurs**, *GE4 award*
American-European Engineering Exchange Student: Master Thesis abroad

Journal publications

Adrienne Pichon, **Iñigo Urteaga**, Lena Mamykina, and Noémie Elhadad. Informing the Design of Individualized Self-Management Regimens from the Human, Data, and Machine Learning Perspectives. *ACM Transactions on Computer Human Interaction*, February 2025.

Vahid Balazadeh, Michael Cooper, David Pellow, Atousa Assadi, Jennifer Bell, Mark Coastworth, Kaivalya Deshpande, Jim Fackler, Gabriel Funingana, (...), **Iñigo Urteaga**, Stephanie Williams, and Rahul G Krishnan. Red Teaming Large Language Models for Healthcare. *arXiv preprint arXiv:2505.00467*, 2025.

Shalmali Joshi, **Iñigo Urteaga**, Wouter A C van Amsterdam, George Hripcsak, Pierre Elias, Benjamin Recht, Noémie Elhadad, James Fackler, Mark P Sendak, Jenna Wiens, Kaivalya Deshpande, Yoav Wald, Madalina Fiterau, Zachary Lipton, Daniel Malinsky, Madhur Nayan, Hongseok Namkoong, Soojin Park, Julia E Vogt, and Rajesh Ranganath. AI as an intervention: improving clinical outcomes relies on a causal approach to AI development and validation. *Journal of the American Medical Informatics Association*, 01 2025.

Iñigo Urteaga and Chris H. Wiggins. Sequential Monte Carlo bandits. *Foundations of Data Science*, 7(4):963–1019, 2025.

Mert Ketenci, **Iñigo Urteaga**, Victor Alfonso Rodriguez, Noémie Elhadad, and Adler Perotte. Variational Shapley Network: A Probabilistic Approach to Self-Explaining Shapley values with Uncertainty Quantification. *arXiv preprint arXiv:2402.04211*, 2024.

Mert Ketenci, Adler Perotte, Noémie Elhadad, and **Iñigo Urteaga**. A Coreset-based, Tempered Variational Posterior for Accurate and Scalable Stochastic Gaussian Process Inference. *arXiv preprint arXiv:2311.01409*, 2023.

Iñigo Urteaga and Chris H. Wiggins. Nonparametric Gaussian mixture models for the multi-armed contextual bandit. *arXiv preprint arXiv:1808.02932*, 2023.

Iñigo Urteaga, Sharon Lipsky-Gorman, Mollie McKillop, and Noémie Elhadad. User Engagement Metrics and Patterns in Phendo, an Endometriosis Research Mobile App. *Nature Partner Journal Digital Medicine*, 2022. (Under review, Minor revisions.).

Kathy Li, **Iñigo Urteaga**, Amanda Shea, Virginia J. Vitzthum, Chris H. Wiggins, and Noémie Elhadad. A predictive model for next cycle start date that accounts for adherence in menstrual self-tracking. *Journal of the American Medical Informatics Association*, 29(1):3 – 11, 09 2021.

Kathy Li, **Iñigo Urteaga**, Chris H. Wiggins, Anna Druet, Amanda Shea, Virginia J. Vitzthum, and Noémie Elhadad. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile health data. *Nature Partner Journal Digital Medicine*, 3(79), 2020.

Iñigo Urteaga, Mollie McKillop, and Noémie Elhadad. Learning endometriosis phenotypes from patient-generated data. *Nature Partner Journal Digital Medicine*, 3(88), 2020.

Iñigo Urteaga and Chris H. Wiggins. (Sequential) Importance Sampling Bandits. *arXiv e-print:1808.02933*, August 2018.

Iñigo Urteaga, Mónica F. Bugallo, and Petar M. Djurić. Sequential Monte Carlo for inference of latent ARMA time-series with innovations correlated in time. *EURASIP Journal on Advances in Signal Processing*, 2017(1), Dec 2017.

Iñigo Urteaga and Chris H. Wiggins. Bayesian bandits: balancing the exploration-exploitation tradeoff via double sampling. *arXiv eprint:1709.03162*, September 2017.

Iñigo Urteaga and Petar M. Djurić. Sequential Estimation of Hidden ARMA Processes by Particle Filtering - Part II. *IEEE Transactions on Signal Processing*, 65(2):494–504., 2016.

Iñigo Urteaga and Petar M. Djurić. Sequential Estimation of Hidden ARMA Processes by Particle Filtering - Part I. *IEEE Transactions on Signal Processing*, 65(2):482–493, 2016.

José María Cabero, **Iñigo Urteaga**, Virginia Molina, Fidel Liberal, and José Luis Martín. Reliability of Bluetooth-based connectivity traces for the characterization of human interaction. *Ad Hoc Networks*, 24, Part A(0):135 – 146, 2015.

José María Cabero, Virginia Molina, **Iñigo Urteaga**, Fidel Liberal, and José Luis Martín. Acquisition of human traces with Bluetooth technology: Challenges and proposals. *Ad Hoc Networks*, 12(0):2–16, 2014.

Iñigo Urteaga, Na Yu, Nicholas Hubbell, and Qi Han. AWARE: Activity aware maintenance of communication structures for wireless sensor networks. *Pervasive and Mobile Computing*, 13:111–124, 2014.

Iraide Unanue, **Iñigo Urteaga**, Ronaldo Husemann, Javier Del Ser, Valter Roesler, Aitor Rodriguez, and Pedro Sanchez. *A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and Performance*, pages 9 – 15. Recent Advances on Video Coding. InTech, 2011.

Kevin Barnhart, **Iñigo Urteaga**, Qi Han, Anura P.Jayasumana, and Tissa Illangasekare. On Integrating Groundwater Transport Models with Wireless Sensor Networks. *Journal of Ground Water*, 48(5), October 2010.

Iñigo Urteaga, Kevin Barnhart, and Qi Han. REDFLAG: A REal-time, Distributed, Flexible, Lightweight, And Generous Fault Detection Service for Data-driven Sensor Applications. *Pervasive and Mobile Computing (PMC) Journal*, 5(5), October 2009.

Peer-reviewed conference proceedings

Eliezer da Silva, Arto Klami, Diego Mesquita, and **Iñigo Urteaga**. On the Identifiability of Tensor Ranks via Prior Predictive Matching. In *Proceedings of the Twenty-Ninth Annual Conference on Artificial Intelligence and Statistics*. PMLR, 02–05 May 2026.

Mert Ketenci, Adler J Perotte, Noémie Elhadad, and **Iñigo Urteaga**. Accurate and Scalable Stochastic Gaussian Process Regression via Learnable Coreset-based Variational Inference. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 2101–2142. PMLR, 21–25 Jul 2025.

Lucia Filippozzi, Santiago Mazuelas, and **Iñigo Urteaga**. Minimax Risk Classifiers for Mislabeled Data: a Study on Patient Outcome Prediction Tasks. In *Proceedings of the 10th Machine Learning for Healthcare*, volume 252 of *Proceedings of Machine Learning Research*, pages 1–52. PMLR, 16–17 Aug 2024.

Iñigo Urteaga, Moulay-Zaïdane Draïdia, Tomer Lancewicki, and Shahram Khadivi. Multi-armed bandits for resource efficient, online optimization of language model pre-training: the use case of dynamic masking. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10609–10627, Toronto, Canada, July 2023. Association for Computational Linguistics.

Iñigo Urteaga, Kathy Li, Chris Wiggins, and Noémie Elhadad. A Generative Modeling Approach to Calibrated Predictions: A Use Case on Menstrual Cycle Length Prediction. In Ken Jung, Serena Yeung, Mark Sendak, Michael Sjoding, and Rajesh Ranganath, editors, *Proceedings of the 6th Machine Learning for Healthcare Conference*, volume 149 of *Proceedings of Machine Learning Research*, pages 535–566. PMLR, 06–07 Aug 2021.

Iñigo Urteaga, Tristan Bertin, Theresa M. Hardy, David J. Albers, and Noémie Elhadad. Multi-Task Gaussian Processes and Dilated Convolutional Networks for Reconstruction of Reproductive Hormonal Dynamics. In *Proceedings of the 4th Machine Learning for Healthcare*, volume 106 of *Proceedings of Machine Learning Research*, pages 66–90. PMLR, 09–10 Aug 2019.

Iñigo Urteaga, Mollie McKillop, Sharon Lipsky-Gorman, and Noémie Elhadad. Phenotyping Endometriosis through Mixed Membership Models of Self-Tracking Data. In *2018 Machine Learning for Healthcare (MLHC)*, 2018.

Iñigo Urteaga and Chris Wiggins. Variational inference for the multi-armed contextual bandit. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 698–706. PMLR, 09–11 Apr 2018.

Iñigo Urteaga and Petar M Djurić. Multiple Particle Filtering for Inference in the presence of state correlation of unknown mixing parameters. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3849–3853, 2017.

Iñigo Urteaga, Mónica F. Bugallo, and Petar M Djurić. Sequential Monte Carlo methods under model uncertainty. In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, pages 1–5, June 2016.

Iñigo Urteaga, Mónica F. Bugallo, and Petar M Djurić. Sequential Monte Carlo sampling for correlated latent long-memory time-series. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6580–6584, March 2016.

Iñigo Urteaga and Petar M Djurić. Particle filtering of ARMA processes of unknown order and parameters. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4105–4109, April 2015.

Susana Pérez-Sánchez, José María Cabero, and **Iñigo Urteaga**. DTN Routing Optimised by Human Routines: The HURRy Protocol. In *Wired/Wireless Internet Communications*, volume 9071 of *Lecture Notes in Computer Science*, pages 299–312. Springer International Publishing, 2015.

Iñigo Urteaga, Mónica F. Bugallo, and Petar M Djurić. Filtering of nonlinear time-series coupled by fractional Gaussian processes. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 489–492, 2015.

Iñigo Urteaga, Mónica F. Bugallo, and Petar M Djurić. Sequential Monte Carlo sampling for systems with fractional Gaussian processes. In *2015 Proceedings of the 23th European Signal Processing Conference (EUSIPCO)*, pages 1246–1250, 2015.

Iñigo Urteaga and Petar M Djurić. Estimation of ARMA state processes by particle filtering. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8033–8037, May 2014.

Douglas E. Johnston, **Iñigo Urteaga**, and Petar M. Djurić. Replication and optimization of hedge fund risk factor exposures. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8712–8716, May 2013.

Iñigo Urteaga, Na Yu, Nicholas Hubbell, and Qi Han. AWARE: Activity AWARE network clustering for wireless sensor networks. In *IEEE Local Computer Networks*, pages 589–596, 2011.

Iñigo Urteaga, Iraide Unanue, Javier Del Ser, Pedro J. Sánchez, and Aitor Rodriguez. On the design of a scalable multimedia streaming system based on receiver-driven flow and congestion awareness. In *2010 International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, pages 39–45, July 2010.

Iñigo Urteaga, Kevin Barnhart, and Qi Han. REDFLAG a Run-timE, Distributed, Flexible, Lightweight, And Generic fault detection service for data-driven wireless sensor applications. In *IEEE International Conference on Pervasive Computing and Communications, 2009*, pages 1–9, March 2009.

Peer-reviewed workshop publications

Changhun Lee, Lena Mamykina, and **Iñigo Urteaga**. Enhancing the Perceived Quality of Task-Oriented Chatbots: Can In-Context Reinforcement Learning be a Breakthrough? In *INFORMS Workshop on Data Science 2025*. INFORMS College on Artificial Intelligence, October 2025.

Iker Pérez, Josu Ceberio Uribe, and **Iñigo Urteaga**. Instance generation for maximum independent set using graph generative models. In *XVI Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados:(MAEB 2025)*, May 2025.

Matthew Levine and **Iñigo Urteaga**. Data-assimilation meets automatic dierentiation for identication of dynamical systems from irregularly-sampled, noisy data. In *XXVIII Congress of Differential Equations and Applications (CEDYA) & XVIII Congress of Applied Mathematics (CMA)*. Spanish Society of Applied Mathematics (SeMA), June 2024. *Oral presentation*.

Iñigo Urteaga, Moulay-Zaïdane Draïdia, Tomer Lancewicki, and Shahram Khadivi. Gaussian Process Thompson sampling for Bayesian optimization of dynamic masking-based language model pre-training. In *NeurIPS 2022 Workshop "Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems"*, December 2022.

Iñigo Urteaga, Moulay-Zaïdane Draïdia, Tomer Lancewicki, and Shahram Khadivi. Thompson sampling for interactive Bayesian optimization of dynamic masking-based language model pre-training. In *EMNLP 2022 Workshop "Novel Ideas in Learning-to-Learn through Interaction" (NILLI)*, December 2022. *Lightning Talk*.

Iñigo Urteaga and Noémie Elhadad. Human-Centered Reinforcement Learning for Personalized Self-Management Strategies. In *CHI 2022 Workshop "Grand Challenges for Personal Informatics and AI"*, May 2022.

Iñigo Urteaga and Chris H. Wiggins. Sequential Monte Carlo for Multi-Armed Bandit Agents. In *5th Workshop on Sequential Monte Carlo Methods*, April 2022.

Kathy Li, **Iñigo Urteaga**, Amanda Shea, Virginia Vitzthum, Chris H Wiggins, and Noémie Elhadad. A generative, predictive model for menstrual cycle lengths that accounts for potential self-tracking artifacts in mobile health data. In *NeurIPS 2020 Workshop "Machine Learning for Mobile Health"*, 2020. *Contributed Talk*.

Kathy Li, **Iñigo Urteaga**, Amanda Shea, Virginia Vitzthum, Chris H Wiggins, and Noémie Elhadad. A generative, predictive model for menstrual cycle lengths that accounts for potential self-tracking artifacts in mobile health data. In *Machine Learning in Science & Engineering (MLSE2020)*, 2020. *Spotlight talk*, Health Sciences track.

Iñigo Urteaga and Chris H. Wiggins. Bandits with sequentially observed rewards: a Bayesian generative Thompson sampling approach. In *NeurIPS 2018 Workshop "Reinforcement Learning under Partial Observability"*, 2018.

Iñigo Urteaga and Chris H. Wiggins. Nonparametric Gaussian mixture models for the multi-armed contextual bandit. In *NeurIPS 2018 Workshop "All of Bayesian Nonparametrics (Especially the Useful Bits)"*, 2018.

Iñigo Urteaga and Chris H. Wiggins. Sequential Monte Carlo for Dynamic Softmax Bandits. In *1st Symposium on Advances in Approximate Bayesian Inference (AABI 2018)*, 2018.

Iñigo Urteaga, David J. Albers, Marija Vlajic Wheeler, Anna Druet, Hans Raffauf, and Noémie Elhadad. Towards Personalized Modeling of the Female Hormonal Cycle: Experiments with Mechanistic Models and Gaussian Processes. In *NeurIPS 2017 Workshop "Machine Learning for Health"*, 2017.

Academic leadership & mentoring

Ph.D. dissertation defense committee member

- *Forecasting Tasks from Electronic Health Record Data: Model Selection, Interpretability, and Scalability* by Mert Ketenci (25/09/2025) Department of Computer Science, Columbia University
- *Predictive Machine Learning for menstrual cycle data* by Kathy Li (04/11/2022) Department of Applied Physics and Applied Mathematics, Columbia University
- *Learning Latent Variable Models: Efficient Algorithms and Applications* by Matteo Ruffini (02/14/2019) Department of Computer Science of Universitat Politecnica de Catalunya

Ph.D. student supervision, advising & mentoring

- Lucia Filippozzi (expected graduation 06/2026), co-advised with Prof. Claudio Agostinelli
  Mathematics, University of Trento
- Mert Ketenci (graduated 09/2025), advised by Prof. Noémie Elhadad
  Computer Science, Columbia University
- Kathy Li (graduated 05/2022), advised by Prof. Chris Wiggins
  Applied Physics and Applied Mathematics and Data Science Institute, Columbia University
- Adrienne Pichon (expected graduation 12/2023), advised by Prof. Noémie Elhadad
  Department of Biomedical Informatics, Columbia University
- Gal Levy-Fix (graduated 06/2020), advised by Prof. Noémie Elhadad
  Department of Biomedical Informatics, Columbia University
- Mollie McKillop (graduated 05/2019), advised by Prof. Noémie Elhadad
  Department of Biomedical Informatics, Columbia University

Master student supervision, advising & mentoring

- Sofia Díez Ocharan (expected graduation 03/2026)
  "Modelado estadístico de datos con patrones de ausencia no aleatorios".
  EHU-UPV Master on Statistical, Computational and Mathematical Modeling
- Aitor Diaz Uriondo (graduated 01/2025)
  Gaussian Process regression: fundamentals and applications.
  EHU-UPV Master on Statistical, Computational and Mathematical Modeling
- Regis Konan Marcel Djaha (graduated 07/2024)
  Understanding Autoencoders and Variational Autoencoders.
  BCAM, African Institute for Mathematical Sciences (AIMS)
- Quentin Chu (graduated 06/2024)
  Reinforcement Learning for individualized self-management using via mobile health.
  Computer Science, Columbia University
- Siddhant Pravin Mahurkar (graduated 12/2022)
  Natural Language Processing and Reinforcement Learning in the context of a nutritional chatbot.
  Data Science Institute Scholars Project, Columbia University
- Moulay Zaidane Al Bahi Draidia (graduated 06/2022)
  Multi-armed bandit optimization for Transformer-based natural language models.
  Data Science Institute, Columbia University
- Kenny Jin (graduated 12/2021)
  Transformer-based natural language models: pre-training and fine-tuning.
  Data Science Institute, Columbia University
- Aimee Moses (graduated 12/2020)
  Statistical signal processing for self-tracked mobile health data.
  Applied Mathematics, Columbia University

Undergraduate student supervision, advising & mentoring

- Iker Zabala (Summer, Fall 2024)
  Graph-generative machine learning, Computer Science and Artificial Intelligence, EHU/UPV.
- Dinko Franceschi (Spring 2018)
  Multi-armed bandits, Data Science Institute, Columbia University.
- Edward Yu (Spring & Fall 2017), co-advised with Prof. Chris Wiggins
  Multi-armed bandits and statistical data analysis, Data Science Institute, Columbia University.
- Su Hang (Fall 2016, Spring 2017), co-advised with Prof. Chris Wiggins
  Statistical data analysis for cancer datasets, Data Science Institute, Columbia University.
- Malvin De Nunez, Jouse Nassar, Ian Jacobsen, William Dwyer (Spring 2016)
  SVMs for fetal heart-rate classification, Senior design project, Stony Brook University
- Lars Folkerts, Shan Liu (Fall 2013 – Spring 2014)
  Summer research, statistical signal processing and webserver development, Stony Brook University.

Scientific presentations

Scientific talks

- "*Hierarchical learning of partially observed interrelated dynamical systems*", Speaker
  2024 SIAM Conference on Mathematics of Data Science (24/10/2024)
- "*Generative, probabilistic machine learning for biomedical applications: incorporating scientific knowledge into data-driven learning.*", Invited Speaker
  CIC bioGUNE - IkerBasque Scientific seminar series: emerging group leaders in Euskadi, (06/06/2024)
- "*Sequential Monte Carlo bandits*",
  Universidad Carlos III - Signal Processing Group scientific seminars, Invited Speaker (24/05/2024)
- "*Probabilistic Machine learning for predictive models in healthcare: a use case on mobile health data*",
  IE-RSME Workshop on Applied Mathematics: AI challenges in Health Care, Invited Speaker (17/05/2024)
- "*A Coreset-based, Tempered Variational Posterior for Accurate and Scalable Stochastic Gaussian Process Inference*", BCAM & EHU Joint Machine Learning Workshop, Invited talk (21/03/2024)
- "*Multi-armed bandits for resource-efficient online optimization of language model pre-training: the use case of dynamic masking*", eBay Inc. Applied Research, Invited talk (01/11/2023)
- "*Probabilistic Machine Learning for Menstrual Cycle Length Predictions via mobile health apps: disentangling menstruation patterns from self-tracking adherence*"
  Applied Center for Data Science Seminar, Western Kentucky University. Invited Speaker (11/04/2022)
- "*Probabilistic machine learning for predictive models of mobile health data*"
  The University of Iowa Computer Science Department Colloquium. Invited Speaker (09/26/2022)
- "*Statistical learning of the menstrual cycle from noisy and missing hormone observations.*"
  Banff International Research Station's workshop "*BIRS Dynamics and Data Assimilation, Physiology and Bioinformatics: Mathematics at the Interface of Theory and Clinical Application*". (06/02/2022)
- "*Statistical Learning Of Menstruation From Indirect, Noisy And Missing Observations.*"
  The Rockefeller University Physics-Biology Center Studies, Seminar Series Invited Speaker (10/19/2021)
- "*Bayesian models and inference for reinforcement learning: multi-armed bandits for practical use.*"
  Corning.Inc Data Science Invited Speaker Series (04/28/2021)
- "*Bayesian models and inference for flexible and efficient multi-armed bandits*"
  eBay Research and University Partnership for Technology Tech Talk series (03/24/2021)
- "*Learning Across a Healthcare Data Network to Improve Model Robustness and Evidence Reliability*"
  Panelist, 2019 American Medical Informatics Association Symposium, Washington, D.C. (11/20/2019)
- "*Bayesian modeling and inference for predictive and prescriptive applications*"
  Basque Center for Applied Mathematics, Scientific Seminar, Bilbao (10/01/2019)
- "*Sequential Monte Carlo Bandits*"
  Multi Armed Bandit Workshop, Imperial College London (09/25/2019)
- "*Bayesian models and inference for reinforcement learning: the multi-armed bandit case*"
  Department of Computer Science of Universitat Politecnica de Catalunya (02/15/2019)
- "*Variational Inference for the Multi-Armed Contextual Bandit*"
  12th Annual Machine Learning Symposium, New York Academy of Sciences (03/09/2018)
- "*The multi-armed bandit: from slot-machines to medicine*"
  Columbia University APAM research conference (10/13/2017)
- "*The multi-armed bandit: from slot-machines to medicine*"
  Columbia University Postdoctoral Seminar Series (09/08/2017)
- "*In Search of the Dynamics of Time-Varying Phenomena*"
  Stony Brook University, Provost Graduate Lecture Series (03/24/2016)

Media presence and Outreach activities

- "*La toma de decisiones: ¿Cuánto afectan y cómo intervienen los sistemas de computación?*
  Speaker at "Pintxo-Pote Zientifikoa" outreach event in Bilbao (2024/01/18)'
- Research within Elhadad's lab featured in Scientific American's video article on endometriosis (12/01/2022):
  "*One in ten people who menstruate suffer from endometriosis: why do we know so little about it?*"
- "*Ciencia de datos y salud (Data science in healthcare)*"
  "*La mecánica del Caracol*" Radio Euskadi (01/24/2022)
- "*Data Science Research: A Bayesian view of multi-armed bandits*"
  Comunidade Data Science Brazil, Invited Speaker (09/09/2021)
- "*Data science for reconstruction and prediction of female reproductive hormones*"
  Data Science Institute Scholars Program Seminar, Columbia University (07/30/2019)
- "*An introduction to the multi-armed bandit problem*"
  Columbia University Summer@SEAS research seminars (07/19/2017)

Other Presentations

- "*Adapting multi-armed bandits to real-life: Flexible models and approximate inference*"
  Columbia Data Science Institute: Foundations of Data Science Center (11/19/2018)
- "*Probabilistic Phenotyping of Endometriosis from Self-Tracking Data*"
  NSF Smart and Connected Health workshop, University of Virginia (09/24/2018)
- "*Sequential Importance Sampling Bandits*"
  Columbia Data Science Day (03/28/2018)
- "*Bayesian bandits: balancing exploitation/exploration tradeoff via double sampling*"
  Columbia Data Science Day (04/05/2017)
- "*Variational inference for the multi-armed contextual bandit problem with linear Gaussian Mixture Models*"
  Frontiers in Computing Systems Symposium, Columbia Data Science Institute (03/24/2017)
- "*EHR Predictive Analytics as a Survival Task*"
  NSF Smart and Connected Health workshop, Boston University (03/21/2017)

Professional service

Journal and Book Editorial Boards and Reviewing

- Editorial board, Action Editor: Transactions on Machine Learning Research (2025-)
- Editorial board of reviewers: Journal of Machine Learning Research (2022-)
- Reviewing: Transactions on Machine Learning Research (TMLR) Expert Reviewer, Journal of Machine Learning Research (JMLR) 2019-2022, Springer Statistics books, Springer Statistics and Computing, IEEE Transactions on Signal Processing (TSP), PLOS ONE, IEEE Signal Processing Letters, Statistics & Probability Letters, EURASIP Journal on Advances in Signal Processing, Communications in Statistics: Simulation and Computation, Digital Signal Processing, Computer Communications, Engineering Optimization

Conference Editorial Boards and Reviewing

- Program Committee member: AISTATS 2025-2026 (Area Chair), MLHC 2023-2026 (Program Chair, General Program Chair, Board), ACML 2022-2025 (Area Chair)
- Organizer: 2024 SIAM Conference on Mathematics of Data Science: "*Data-Driven Learning of Dynamical Systems from Partial Observations*" minisymposium
- Reviewing: ICML 2019-2025 (Expert Reviewer), NeurIPS 2018-2025, AISTATS 2018-2024, IEEE ICASSP 2020-2024 & 2016-2017, MLHC 2018-2023, EUSIPCO 2020, NeurIPS2023 Workshop on Deep Generative Models for Health, UAI 2019, AABI Symposiums 2018-2020 & 2023-2024, AABI NIPS2017 workshop, MAEB 2025, FUSION 2017, IEEE SAM 2016, IEEE CAMSAP 2015 Systems, 2018.

Scientific evaluation committees

- Schmidt Sciences, Advisor: AI 2050 Early Career Fellowship external evaluator, 2025
- Bilateral Research Cooperation Program MOST-FRQS: Artificial Intelligence and healthcare research program evaluator, 2022.
- National Science Foundation: Directorate of Computer & Information Science, Engineering and Division of Information and Intelligent Systems, 2018.

IEEE Signal Processing Society Member

SIAM – Society for Industrial and Applied Mathematics Member

Teaching Experience

2025 – 2026 **Instructor**, *Euskal Herriko Unibertsitatea (EHU)*

- Springs of 2025 & 2026 - *"Applied Artificial Intelligence and its Mathematical Foundations"*
  Graduate course on "Statistics for AI"

2016 – 2018 **Instructor**, *Columbia University*

- Spring 2018 - *Graduate seminar "Data Science for mHealth"*: Classification and clustering for mHealth.
- Fall 2017 - *"Biomedical Informatics Data Mining seminar"*: Introduction to compressed sensing.
- Summer 2017 - *"Biomedical Informatics Data Mining seminar"*: Introduction to Monte Carlo estimation.
- Fall 2016 - *"Biomedical Informatics Data Mining seminar"*: Optimization techniques in deep-learning.

2011 – 2015 **Teaching Assistant**, *Stony Brook University*

- Spring 2015 - *Random Signals and Systems* (ESE 306)
- Fall 2014 - *Introduction to Electrical Engineering* (ESE 123)
- Spring 2014 - *Random Signals and Systems* (ESE 306)
- Spring 2013 - *Random Signals and Systems* (ESE 306)
- Fall 2012 - *Introduction to Electrical Engineering* (ESE 123)
- Spring 2012 - *Introduction to Electrical Engineering* (ESE 123)
- Fall 2011 - *Introduction to Electrical Engineering* (ESE 123)

2012 – 2014 **SBU Engineering Summer Camps**, *Stony Brook University*
Instructor and Teaching Assistant for high-school students interested in engineering careers.

Languages

- Spanish (**Native speaker**)
- English (**Professional level**)
- Euskera (**Native speaker, EGA**)
- French (**Intermediate level**)

Technical skills

OS Linux (Ubuntu, Debian), Windows

Programming Python (NumPy, SciPy, scikit-learn, Tensorflow/PyTorch), Matlab/Octave, Git, Shell scripting