Alexandre Andorra

Machine Learning Scientist

Probabilistic ML, Causal Inference & Model Evaluation

andorra.alexandre@gmail.com LinkedIn GitHub alexandorra.github.io

San Francisco Bay Area US Green Card (EB-1A, Extraordinary Ability) English, French, Spanish



Machine learning scientist with **8+ years** building and evaluating production probabilistic AI systems across tech, pharma and professional sports. I specialize in Bayesian ML, causal inference, time series and rigorous model evaluation. At **Meta**, I measure the causal impact of AI tools on topline metrics across the family of apps. **Core contributor to PyMC** (8K+ GitHub stars), co-founder of **PyMC Labs** ($0 → $1M annual revenue in 2 years), principal modeler on a paper **under review in Nature**, and creator of **Learning Bayesian Statistics**, a top-1.5% global podcast.

Experience

Meta Data Scientist — Apr 2026 – Present, Menlo Park, CA

- Lead causal measurement of the productivity impact and ROI of AI tools — coding assistants and internal agents — on topline metrics across Meta's family of apps.
- Build quasi-experimental and probabilistic models — difference-in-differences, synthetic control, interrupted time series, state space models, Gaussian processes, IRT models — to quantify AI-driven impact at scale.
- Develop interpretable production forecasts and agentic decision support whose outputs are inputs to exec-level AI investment decisions.

Miami Marlins Senior Applied Scientist — Jul 2024 – Oct 2025, Miami, FL

- Built production Bayesian forecasting models on 1M+ tracking events (Trackman, HawkEye) to predict player performance, improving projection accuracy by 15% over baseline.
- Developed hierarchical time series models integrating player tracking, biomechanics and game data to inform $100M+ roster decisions.
- Translated probabilistic model outputs into actionable insights for coaching staff and front office, directly influencing game strategy and player development.
- Led technical education initiatives upskilling the 10-person R&D team in advanced statistical methods.

PyMC Labs Principal Data Scientist & Co-founder — Mar 2021 – Jul 2024, Remote

- Co-founded and scaled the company from $0 to $1M annual revenue in 2 years (50% YoY client growth), securing 15+ enterprise contracts and establishing it as the leading Bayesian ML consultancy.
- Architected production systems: Bayesian A/B testing frameworks, causal inference pipelines (treatment effects, substitution effects, backdoor adjustment), Gaussian process models for spatio-temporal forecasting.
- Delivered for Fortune 500 clients in pharma R&D (RNA vaccine development), agtech (crop optimization) and e-commerce (marketing attribution reaching 10M+ users), leading cross-functional teams.
- Designed and delivered 40+ technical workshops training 500+ practitioners in hierarchical modeling, causal inference, time series and decision optimization.

Earlier: Data Scientist at Société.com (2020–21), 360Learning (2019–20) and Banque de France (2018–19): financial prediction models, ETL pipelines, automated financial reporting.

Research & Open Source

Baygent Skills 2026 – present — GitHub

Open-source agent skills that teach AI coding assistants rigorous Bayesian modeling, causal inference and model evaluation; works with Claude Code, Codex, Cursor and Gemini CLI.

Hidden Diversity of Threatened Sharks and Rays in the Global Meat Trade bioRxiv
Co-author and principal modeler (bespoke Bayesian hierarchical model); under review in Nature.

The Soccer Factor Model 2024 – present arXiv:2412.05911 Dashboard
Hierarchical Bayesian model separating player skill from team effects over time; introduced Skill and Performance Above Replacement; interactive rankings of 1,000+ players with uncertainty.

PyMC Core developer, 2019 – present pymc.io Gallery
Implemented the ZeroSumNormal distribution, improved the multivariate normal API and created prior helper functions. Authored 8 official tutorials (HSGP, multilevel modeling, Kronecker covariances, LKJ priors) used by thousands of practitioners.

PollsPosition Electoral forecasting Blog
First French polling aggregator, built on MRP models, with GP and HMM models of presidential approval; fully open source; 4K visitors/month during campaigns.

Teaching & Community

Learning Bayesian Statistics Creator & host, 2019 – present
Top-1.5% global podcast and the largest Bayesian statistics community: 12K monthly listeners, 165+ episodes, 100+ paid community members. Currently hosting a 12-episode series on reinforcement learning.

Intuitive Bayes Co-founder & instructor, 2021 – present
Online platform teaching probabilistic AI to 200+ students ($100K+ revenue).

Selected Talks

- Was It the Player or the Team? Building a Bayesian Skill Model for Soccer — Carnegie Mellon University, live workshop, Oct 2026
- What Your Mixed Model Already Believes — Visiting Speaker Seminar, Princeton University, Sep 2026
- Amortized Inference & GPUs for Stan — live at StanCon, Aug 2026
- A Beginner's Guide to State Space Modeling — PyData Berlin, 2025
- Mastering Gaussian Processes with PyMC — PyData NYC, 2024
- Live podcast show — StanCon Oxford, 2024
- Guest on Stats + Stories and the Super Data Science podcast

Skills

Model evaluation	Proper scoring rules, calibration (PIT, LOO-PIT), LOO-CV and model comparison, posterior predictive checks, A/B testing and experimentation, sequential analysis
Causal inference	Difference-in-differences, synthetic control, interrupted time series, propensity scoring, backdoor adjustment, DAGs
Machine learning	Bayesian ML, probabilistic AI, hierarchical models, Gaussian processes, state space models, IRT, Bayesian optimization, deep neural networks, AI agents
Tools	Python, PyMC, NumPyro, JAX, PyTorch, TensorFlow, NumPy, pandas, scikit-learn

Education

HEC Paris — MSc, Management
Université Paris-Sorbonne — BSc, History
Freie Universität Berlin — MSc, Public Policy
Causal AI — Robert Osazuwa Ness, Microsoft Research