Othman M. Benomar, PhD — Senior / Principal Data Scientist

Tokyo, Japan • +81 70 3231 9836 • othman.benomar.pro@gmail.com
GitHub: github.com/OthmanB • Website: othmanbenomar.dev • LinkedIn: othman-benomar

Professional Summary

Lead Data Scientist with 15+ years of experience designing large-scale ML systems, advanced statistical models, and data-driven decision frameworks in industry and space physics. Expert in Bayesian statistics, deep learning, time-series modeling, anomaly detection, and ML infrastructure. Background in asteroseismology and physical-system inversion, with strong cross-domain transferability. Methodical first-principles thinker and builder of state-of-the-art scientific models from the ground up.

Core Competencies

- ML: Transformers, autoencoders, CNN/LSTM, retrieval, embeddings (applied in production and research contexts), Bayesian inference, problem optimization.
- AI/GenAI: RAG pipelines, LLM orchestration (design), vector databases, embeddings.
- Big Data: Python, SQL, distributed computing, ETL pipelines.
- Engineering: Kubernetes, Docker, MLFlow, Prometheus/Grafana, HPC workflows, AWS.
- Experimentation: A/B tests, causal inference, modeling, metric design.
- Leadership: mentoring, international project direction.

Leadership & System Ownership

- Led design decisions for mission-critical analytical systems used by the ESA space agency.
- Chaired and contributed to technical workshops defining data standards, validation strategies, and failure modes.
- Took responsibility for architectural trade-offs across accuracy, compute cost, robustness, and interpretability.
- Mentored researchers and engineers across multiple institutions, shaping long-term technical direction.

Professional Experience

Lead Data Scientist — Craftsman Software, Tokyo, Japan (May 2024–Present)

- Designed and deployed real-time ML systems for anomaly detection in distributed cloud environments.
- Built CNN–LSTM and attention-based architectures for sequential modeling and prediction.
- Built Bayesian adaptive statistical log-anomaly detector based on execution path patterns revealed by Drain3.
- Implemented RAG + MCP + LLM pipelines for context-aware log and code analysis.
- Developed scalable MLOps workflows: feature stores, model governance, inference monitoring.
- Collaborated with engineering leadership on product decisions and long-term ML strategy.

Project Associate Professor (Postdoc) — National Astronomical Observatory of Japan, Tokyo (Jul 2019–June 2024)

- Directed statistical and ML research programs for multi-terabyte sequential datasets in HPC environments.
- Developed forecasting, classification, anomaly detection, and inversion models.
- Created data validation, interpretability, and uncertainty quantification pipelines.
- Supervised multiple PhD students and built internal ML training curriculum.
- Led international collaborations with NASA, ESA, and research institutions.

Research Associate — New York University Abu Dhabi, UAE (November 2015–June 2019)

- Built large-scale statistical and ML pipelines for sequential and imaging space satellite data.
- Developed physical models analogous to ranking/retrieval and personalization tasks.
- Collaborated with UAE stakeholders on data-driven research initiatives.
- Supervised multiple PhD students on time series pipeline-related analysis projects.

Research Fellow — University of Tokyo (JSPS), Japan (October 2013–October 2015)

- Developed Bayesian inversion and latent-variable models for complex physical systems.
- Secured ¥11M competitive grant
- Taught advanced statistical modeling for astrophysics to graduate students.

Postdoctoral Fellow — University of Sydney, Australia (October 2010–October 2013)

- Built high-dimensional Bayesian models for space-telescope telemetry.
- Secured ARC funding; supervised research students.
- Taught physics to undergraduate students.

Education

PhD — Applied Physics: Statistical Methods Applied to Stellar Plasma Physics (Asteroseismology), Université Paris-Sud XI & École Polytechnique

MSc — Plasma & Optics Physics, Université Paris-Sud XI & École Polytechnique

Selected Technical Achievements

- Developed ML systems processing millions of sequential events per hour with real-time constraints.
- Designed inversion models structurally analogous to ranking, retrieval, and representation learning.
- Architected distributed training pipelines using HPC and cluster compute.
- Delivered complex, multi-stakeholder analytical results with up to 20+ contributors, coordinating modeling choices, validation, and final delivery under strict peer review.

Tools & Technologies

Python, SQL, C++, Go, R, PyTorch, TensorFlow, Keras, MLFlow, HuggingFace, Docker, Kubernetes, Spark, Prometheus, Grafana, PostgreSQL, MySQL, NoSQL, Plotly, Dash, Matplotlib.

Languages: French (native), English (fluent), Japanese, Moroccan Arabic (conversational)

Individual interests

- Blockchain technology enthusiast since 2017: Developed pipelines for collecting and interpreting cryptocurrency market data.
- Co-owner of a restaurant in Sydney, Australia since 2021: Perform management and analytics for staffing and sales.