Sebastián Bobadilla Suárez

VP, Data Science · Applied ML, Production Agentic AI & Knowledge Graph Platforms

UK (Remote) | sebastian.bobadilla.s@gmail.com | bobaseb.github.io | github.com/bobaseb | in/sebastian-bobadilla-suarez

SUMMARY

VP of Data Science moving agentic AI from prototype to production in regulated financial services. Started and lead the company's ontology and knowledge-graph programme, production anomaly-detection agents, contract-first agent APIs, and a company-wide AI-coding transformation. Promoted twice in 3.5 years from Director to VP; background as a PhD computational cognitive neuroscientist (UCL) and Turing Institute Enrichment Fellow.

EXPERIENCE

OnCorps — AI agents for fund-operations oversight in financial services — UK (Remote)

VP, Data Science (from Feb 2026) · VP, Machine Learning (Dec 2025 – Feb 2026) — Dec 2025 – present

- **Ontology & Knowledge Graph Programme:** Started and drive the company's financial-services ontology programme (OWL/RDF/FIBO, Apache Jena Fuseki/SPARQL), authoring 95% of its human-written commits; exposed to product agents via read-only MCP with per-client data segregation and rule provenance. A dedicated ontology squad is now being formed, with this role as technical leader.
- **Production Anomaly-Detection Agents:** For a global fixed-income asset manager (∼$2.3 trillion AUM), evolved daily fund income-projection anomaly detection from MAD-score statistics to an agentic investigation layer. Onboarded 5 federated Snowflake feeds via Unity Catalog Lakehouse Federation, built 15 rule-based and 23 statistical checks, and encoded analyst playbooks as agent skills. Audited and fixed production defects (sign error in 107 of 121 exports, replay non-determinism, stale comparisons), freed 1.61 TB of leaked storage, and added fail-closed guards.
- **Contract-First Agent Platform:** Tech lead for self-service check authoring across 2 products. Designed a unified OpenAPI 3.1 agent platform with semver breaking-change gates, generated Python/TypeScript SDKs, offline evaluation suites (golden sets, deterministic scorers, LLM judges), and fail-closed security scans blocking destructive SQL and external URLs.
- **Always-On Multi-Agent POC (Legion):** Built an internal proof-of-concept always-on agent and alternative to Claude Code (not in production; currently in personal work use and championed for wider colleague adoption) for multi-week memory and multi-repo orchestration. Designed a hierarchical memory tree compressing multi-week conversations into 128 KB without deleting history, an orchestrator–manager–worker crew driving Claude Code across repositories, and web/TUI/iOS/chat interfaces; hardened via 4 security review rounds and mutation testing (1,095 passing tests).
- **Agentic Coding Transformation:** Led a company-wide agentic-coding programme for 2 months (Apr–Jun 2026; 20 epics across 2 tracks), then handed over. Trained all engineers, pushed security hooks and Ruff security lint gates, set a $200/day AI spend cap, launched a shared agent-skills plugin marketplace, and reported weekly to the ELT and board.
- **Delivery Telemetry & Empirical Research:** Analysed 6 months of engineering telemetry across 442 repositories, identifying distinct per-product bottlenecks (e.g., 27.9-day release waits vs. 40.6-day work-item lifetimes) to deliver 4 targeted CEO decisions. First author of *Audit the Scaffold, Not the Checkpoint* (arXiv:2609.34924), combining Lean 4 proofs, SWE-bench experiments, and multilevel modelling of 378 production agentic sessions.
- **Leadership, Governance & Agent Orchestration:** Manage 2 direct reports; interviewed and supported 2 data-science hires; reviewed 385 PRs in 2026; member of the ISO 42001 AIMS management board. Orchestrated parallel coding agents (600+ merged PRs across 19 repos in 2026) through human review gates and Red Zone security sign-off.

Sr. Director, Machine Learning (from Apr 2024) · Director, Machine Learning — Apr 2022 – Dec 2025

- **Explainable Anomaly Detection:** Designed and shipped massively univariate MAD-score anomaly detection on Databricks for a ∼$2.3 trillion AUM fixed-income asset manager, ranking daily fund income-projection line items against 3-month histories and multiple temporal deltas in parallel.
- **Production Dashboards & Stack Modernisation:** Owned 4 production pricing dashboards for two Fortune 500 IT services firms; migrated 10,000+ LOC from R Shiny to Python Dash on a custom component library, modernised ETL and serving from Pandas/SQLite to Polars, Python 3.12, Dash 3, and AWS RDS with query caching, and deployed via Kubernetes.
- **Profit Modelling & Applied AI Systems:** Built a deal profitability model for the same two Fortune 500 IT services firms ($R^2 = 0.935$ and > 0.8), outperforming prior modelling efforts and re-engaging a key client. Built RAG and agent chatbots (AWS Bedrock, OpenAI, Qdrant/Databricks Vector Search, Langfuse/MLflow tracing, MCP) for a ∼$21B-revenue IT services firm, a SQL-agent prototype for one of the world's largest asset managers (∼$13 trillion AUM), and a 94%-accuracy SBERT classifier.
- **AI Tooling Standardisation:** Piloted and ran the risk assessment that established Claude Code and Cursor as standard AI coding tools; trained engineers and non-technical executives.

University College London (UCL) London, UK

Research Associate in Computational Modelling of Brain Imaging Data Jan 2018 – Nov 2021

- Applied statistical learning, Bayesian/regularized regression, and deep learning to terabyte-scale neuroimaging (fMRI/intracranial) and behavioural datasets with Prof. Brad Love; published in *Nature Communications*, *Nature*, *Communications Biology*, and *Cognitive Psychology*, and supervised 5 graduate/undergraduate ML theses.

The Alan Turing Institute London, UK

Enrichment Year Fellow & Research Intern Jun 2016 – Sep 2017

- Led an interdisciplinary team of statisticians and computer scientists evaluating classifier-induced neural similarity spaces, resulting in a peer-reviewed publication (*Computational Brain & Behavior*); organised and moderated the institute's PhD Topical Discussions series.

EDUCATION

PhD in Cognitive Neuroscience · University College London (UCL), London, UK 2013 – 2018

Decision heuristics, belief updating, delegation, and computational neuroimaging (Profs. Brad Love & Tali Sharot). CONACYT Scholar.

Enrichment Year Fellow · The Alan Turing Institute, London, UK 2016 – 2017

MSc in Social Cognition (Distinction) · University College London (UCL), London, UK 2012 – 2013

BA in Social Anthropology (Honours, GPA 9.7/10) · Escuela Nacional de Antropología e Historia (ENAH), Mexico City 2007 – 2011

SELECTED PUBLICATIONS (14 ARTICLES & PREPRINTS)

Bobadilla-Suarez, S., Suh, B., & Fortin, R. (2026). *Audit the Scaffold, Not the Checkpoint: A Stationarity Dichotomy for Recursive Self-Improvement in Agentic Coding*. arXiv:2609.34924 [cs.LG] (first author; Lean 4 proofs, SWE-bench & production-session study).

Suh, B., Fortin, R., & **Bobadilla-Suarez, S.** (2026). How "Deep Industry Research Agents" Can Change Your Organization. *Harvard Business Review*, March 2026. hbr.org.

Zhang, X. Y., **Bobadilla-Suarez, S.**, … & Love, B. C. (2025). Adaptive stretching of representations across brain regions and deep learning model layers. *Nature Communications*. doi:10.1038/s41467-025-65231-y.

Bobadilla-Suarez, S., Jones, M., & Love, B. C. (2022). Robust priors for regularized regression. *Cognitive Psychology*, 132, 101444. doi:10.1016/j.cogpsych.2021.101444.

Additional peer-reviewed work in **Nature** (2020), **Communications Biology** (2020), **Journal of Neuroscience** (2017), **Journal of Risk and Uncertainty** (2017), *and* **Cornell Law Review** (2017) — *full list at bobaseb.github.io*.

SKILLS

LLMs & Agentic AI: Claude Code, LangGraph-style agent harnesses, agent skills, MCP servers/clients, RAG (Qdrant, Databricks Vector Search, sqlite-vec), AWS Bedrock, LiteLLM, OpenAI, Llama/Mistral, Langfuse, MLflow, LLM evaluation (golden sets, deterministic scorers, LLM judges)

ML & Statistics: Robust statistics (MAD), anomaly detection, regularized regression, multilevel/mixed-effects models, sentence-embedding classifiers (SBERT), Bayesian inference, experimental design, PyTorch, scikit-learn

Data & Platform: Databricks (Asset Bundles, Unity Catalog, Lakehouse Federation, Jobs, DLT, SQL), PySpark, Snowflake, Polars, Pandas, PostgreSQL, SQLite, AWS RDS, ElasticSearch

Knowledge Graphs: OWL, RDF/Turtle, SPARQL, Apache Jena Fuseki, FIBO

Software Engineering: Python, TypeScript, R, SQL, Bash, OpenAPI 3.1 contract-first design, semver gates, SDK codegen, GitHub Actions CI, Docker, Kubernetes, FastAPI, Dash, R Shiny, mutation testing, Lean 4 (formal proofs)

AI Governance & Leadership: ISO 42001 AIMS, ISMS tool reviews, Red Zone human-in-the-loop controls, fail-closed design, programme leadership, ELT/board reporting, hiring & mentoring

Languages: English (fluent), Spanish (native); basic familiarity with Italian, French, and Portuguese