

MADSON ALLAN DE LUNA ARAGÃO

Belo Horizonte, MG | madsondeluna@gmail.com | github.com/madsondeluna | linkedin.com/in/madsonaragao

EXECUTIVE SUMMARY

Bioinformatics Scientist, Machine Learning Analyst, and Software Engineer with 8+ years developing high-performance computational pipelines and data products for production environments. Specialized in building end-to-end ML systems that process large-scale biological datasets deployed on HPC clusters. Combines hands-on technical execution with product ownership expertise (4 years managing complete product lifecycles from ideation to production launch). Proven track record of innovation with 4 filed patents (3 granted), 6 peer-reviewed publications, and multiple international awards. Core competencies include ML/AI engineering, HPC infrastructure optimization (Slurm, PBS, SGE), script engineering for distributed systems, and cross-functional technical leadership.

TECHNICAL STACK

Languages	Python, SQL (PostgreSQL, MySQL), R, Bash, Shell Scripting
HPC & Infrastructure	8+ years: Slurm, PBS, SGE, Job Scheduling, Resource Management, Cluster Optimization, Script Engineering, Parallel Computing, Distributed Systems
ML/AI	Scikit-learn, PyTorch, TensorFlow, XGBoost, Random Forest, SVM, Deep Learning
Product Management	Product Ownership, Lean Inception, OKRs, Roadmapping, Backlog Management, User Research, Metrics-Driven Development, Stakeholder Management
Data/Analytics	Pandas, NumPy, Spark, Matplotlib, Plotly, Statistical Modeling
DevOps & Cloud	Docker, Kubernetes, Git, CI/CD, Nextflow, Snakemake, AWS, GCP, Azure
Agile/Leadership	Scrum, Kanban, Technical Leadership, Cross-functional Teams, Technical Strategy

PROFESSIONAL EXPERIENCE

Head of Bioinformatics & Development <i>Biotech Business</i>	Nov 2025 – Present Brazil
– Technical and product leadership of cross-functional team delivering ML-powered platform for agroindustry – Own product strategy and roadmap: define vision, prioritize features, align stakeholders across R&D and commercial – Architected automated strain identification system reducing manual analysis time from 48h to 2h (Computer Vision + Random Forest) – Built production API serving 50K+ biological records with real-time analytics (FastAPI + PostgreSQL) – Manage product metrics, user feedback loops, and continuous improvement cycles	
Nextflow Ambassador <i>Seqera</i>	Jan 2026 – Present
– Technical advocacy for workflow orchestration in scientific computing (HPC/Cloud) – Community engagement: workshops, technical talks, code reviews	
Machine Learning Researcher – PhD Student <i>Federal University of Minas Gerais (UFMG)</i>	Aug 2024 – Present
– Developing Transformer models for peptide sequence generation (500K+ training samples, 20 species) – Production Nextflow pipelines processing 1M+ sequences/day on HPC infrastructure – Multi-omics data integration using dimensionality reduction (PCA, t-SNE, UMAP)	
Machine Learning Engineer – Master’s Student <i>Federal University of Pernambuco (UFPE)</i>	Jun 2022 – Jul 2024 Recife
– AMPIdentifier (Patent BR 51 2025 005859-4): End-to-end ML pipeline for antimicrobial peptide discovery – Model performance: 96% accuracy, 0.94 F1, 0.98 AUC-ROC (Random Forest, 13,246 samples, 9 engineered features) – Validated discoveries: 47 novel peptides confirmed via structural modeling (AlphaFold2) and MD simulations	

- Open-source adoption: 150+ GitHub stars, cited in 3+ research publications
- RNA-Seq analysis pipeline: differential expression (edgeR/DESeq2), clustering, 15K genes analyzed

R&D Engineer & Product Owner

Oct 2020 – Aug 2023

PickCells

Recife

- **Product Ownership:** Owned 3 products (software/ML/hardware) from ideation to production launch
- Led 5+ Lean Inception workshops with enterprise clients, translating business needs into technical roadmaps
- Managed complete product lifecycle: discovery → prototyping → MVP → iteration → scale
- Defined and tracked OKRs, maintained product backlog, prioritized features based on data and user feedback
- Stakeholder management: C-level executives, technical teams, end users across multiple organizations
- Reduced time-to-market from 6 to 3 months through rapid experimentation and validated learning
- **Technical Delivery:** Computer Vision system for clinical images (92% accuracy, ResNet50, 10K dataset)
- Built IoT platform with real-time analytics: FastAPI backend, React dashboard, WebSocket streaming

SELECTED PROJECTS

decryptAMP | Patent PC202600296 2026

- Automated proteome mining pipeline: enzymatic digestion → feature engineering → ML classification
- Performance: Complete human proteome (20K proteins) in 45min on 8-core system

AMPIdentifier | Patent BR 512025005859-4 2025

- Production ML system: Random Forest (100 estimators), 96% accuracy, 0.94 F1, 0.98 AUC-ROC
- Impact: 47 validated discoveries, 150+ GitHub stars, 3+ citations

EDUCATION

PhD, Bioinformatics 2024 – Present

Federal University of Minas Gerais (UFMG)

MBA, Software Engineering 2024 – Present

University of São Paulo (USP)

MSc, Genetics & Molecular Biology 2022 – 2024

Federal University of Pernambuco (UFPE) | *Ranked 1st/100+ applicants*

Specialization, Data Science & Analytics 2024 – 2026

Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

BSc, Biomedical Sciences 2017 – 2022

Federal University of Pernambuco (UFPE)

PUBLICATIONS & PATENTS

Publications: 6 peer-reviewed papers in Agronomy, Viruses, Crop Science, IJBM, Biopolymers (2022-2026)

Patents (4):

- AMPIdentifier (Software, BR 512025005859-4, 2025) - ML toolkit for antimicrobial peptide prediction
- decryptAMP (Software, PC202600296, 2026) - Bioinformatics tool for encrypted peptide discovery
- SARS-CoV-2 Detection Kit (BR1020210258438, 2021) - Oligonucleotides and method for gene detection
- IoT Microscope (BR10202100250, 2021) - IoT technology for image/video acquisition

Awards & Honors:

- New Generation Bioinformatician Award, X-Meeting 2025 (Brazilian Association for Bioinformatics)
- Young Geneticist of Northeast Brazil, Brazilian Genetics Society (SBG) 2016
- Best Poster Award, XIII Graduate Program Symposium in Genetics & Molecular Biology, UFPE 2023
- 2nd Place (Forensic Genetics), XXI Northeast Genetics Meeting (ENGINE) 2016

Travel Grants:

- ICTP-SAIR School on Biological Physics & Biomolecular Simulations in ML Era 2025
- AI for Protein Design (AI4PD) Workshop, Puerto Varas, Chile 2023
- The Protein Society International Symposium (multiple years)

Other Recognition: Certificate of Excellence in Peer Reviewing (Elsevier 2024) | Honorable Mention in Human & Forensic Genetics (UFPE Research Congress 2017) | Honorable Mention in Chemistry Olympiad (Pernambuco State 2013)

CERTIFICATIONS

AI/ML: Generative AI (MIT) | Oracle Cloud AI | Azure AI & ML (Microsoft) | Advanced Gemini (Google) | ML & Analytics (PUC-Rio) | Advanced Data Analysis (PUC-Rio)
Infrastructure: Slurm Workload Manager (LNCC) | Nextflow Training (Seqera) | GitHub Professional
Product/PM: Data-Driven PM (NASBA) | PM Foundations (PMI) | Requirements Engineering (PUC-Rio) | Product Ownership (Udemy) | Agile PM (Atlassian)

LANGUAGES

Portuguese (Native) | English (Advanced)