

MADSON ALLAN DE LUNA ARAGÃO

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

RESUMO EXECUTIVO

Bioinformatics Scientist, Machine Learning Analyst e Engenheiro de Software com 8+ anos desenvolvendo pipelines computacionais de alto desempenho e produtos de dados para ambientes de produção. Especializado na construção de sistemas ML end-to-end que processam datasets biológicos em larga escala em clusters HPC. Combina execução técnica hands-on com expertise em product ownership (4 anos gerenciando ciclos completos de produto desde ideação até lançamento em produção). Track record comprovado de inovação com 4 patentes depositadas (3 concedidas), 6 publicações peer-reviewed e múltiplos prêmios internacionais. Competências centrais incluem engenharia ML/AI, otimização de infraestrutura HPC (Slurm, PBS, SGE), script engineering para sistemas distribuídos e liderança técnica cross-funcional.

STACK TÉCNICO

Linguagens	Python, SQL (PostgreSQL, MySQL), R, Bash, Shell Scripting
HPC & Infraestrutura	8+ anos: 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, Gestão de Stakeholders
Data/Analytics	Pandas, NumPy, Spark, Matplotlib, Plotly, Modelagem Estatística
DevOps & Cloud	Docker, Kubernetes, Git, CI/CD, Nextflow, Snakemake, AWS, GCP, Azure
Agile/Liderança	Scrum, Kanban, Liderança Técnica, Times Cross-funcionais, Estratégia Técnica

EXPERIÊNCIA PROFISSIONAL

Head de Bioinformática & Desenvolvimento Nov 2025 – Presente
Biotech Business Brasil

- Liderança técnica e de produto de time cross-funcional entregando plataforma ML para agroindústria
- Ownership de estratégia de produto e roadmap: definição de visão, priorização de features, alinhamento de stakeholders em P&D e comercial
- Arquiteteei sistema automatizado de identificação de cepas reduzindo tempo de análise manual de 48h para 2h (Computer Vision + Random Forest)
- Construí API de produção servindo 50K+ registros biológicos com analytics em tempo real (FastAPI + PostgreSQL)
- Gerencio métricas de produto, loops de feedback de usuários e ciclos de melhoria contínua

Nextflow Ambassador Jan 2026 – Presente
Seqera

- Advocacy técnico para orquestração de workflows em computação científica (HPC/Cloud)
- Engajamento com comunidade: workshops, talks técnicos, code reviews

Machine Learning Researcher – Doutorando Ago 2024 – Presente
Universidade Federal de Minas Gerais (UFMG)

- Desenvolvimento de modelos Transformer para geração de sequências peptídicas (500K+ amostras de treino, 20 espécies)
- Pipelines Nextflow de produção processando 1M+ sequências/dia em infraestrutura HPC
- Integração de dados multi-ômicos usando redução de dimensionalidade (PCA, t-SNE, UMAP)

Machine Learning Engineer – Mestrando Jun 2022 – Jul 2024
Universidade Federal de Pernambuco (UFPE) Recife

- AMPIdentifier (Patente BR 51 2025 005859-4):** Pipeline ML end-to-end para descoberta de peptídeos antimicrobianos
- Performance do modelo: 96% accuracy, 0.94 F1, 0.98 AUC-ROC (Random Forest, 13.246 amostras, 9 features engineered)

- Descobertas validadas: 47 novos peptídeos confirmados via modelagem estrutural (AlphaFold2) e simulações MD
- Adoção open-source: 150+ stars GitHub, citado em 3+ publicações de pesquisa
- Pipeline de análise RNA-Seq: expressão diferencial (edgeR/DESeq2), clustering, 15K genes analisados

Engenheiro de P&D & Product Owner

Out 2020 – Ago 2023

PickCells

Recife

- **Product Ownership:** Ownership de 3 produtos (software/ML/hardware) desde ideação até lançamento em produção
- Liderei 5+ workshops Lean Inception com clientes enterprise, traduzindo necessidades de negócio em roadmaps técnicos
- Gerenciei ciclo completo de produto: discovery → prototipagem → MVP → iteração → escala
- Defini e trackeei OKRs, mantive backlog de produto, priorizei features baseado em dados e feedback de usuários
- Gestão de stakeholders: C-level executives, times técnicos, usuários finais em múltiplas organizações
- Reduzi time-to-market de 6 para 3 meses através de experimentação rápida e validated learning
- **Entrega Técnica:** Sistema Computer Vision para imagens clínicas (92% accuracy, ResNet50, dataset 10K)
- Construí plataforma IoT com analytics em tempo real: backend FastAPI, dashboard React, streaming WebSocket

PROJETOS SELECIONADOS

decryptAMP | Patente PC202600296

2026

- Pipeline automatizado de mineração de proteomas: digestão enzimática → feature engineering → classificação ML
- Performance: Proteoma humano completo (20K proteínas) em 45min em sistema 8-core

AMPidentifier | Patente BR 512025005859-4

2025

- Sistema ML de produção: Random Forest (100 estimadores), 96% accuracy, 0.94 F1, 0.98 AUC-ROC
- Impacto: 47 descobertas validadas, 150+ stars GitHub, 3+ citações

FORMAÇÃO

PhD, Bioinformática

Universidade Federal de Minas Gerais (UFMG)

2024 – Presente

MBA, Engenharia de Software

Universidade de São Paulo (USP)

2024 – Presente

MSc, Genética & Biologia Molecular

Universidade Federal de Pernambuco (UFPE) | *Classificado 1º/100+ candidatos*

2022 – 2024

Especialização, Ciência de Dados & Analytics

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

2024 – 2026

BSc, Ciências Biomédicas

Universidade Federal de Pernambuco (UFPE)

2017 – 2022

PUBLICAÇÕES & PATENTES

Publicações: 6 artigos peer-reviewed em Agronomy, Viruses, Crop Science, IJBM, Biopolymers (2022-2026)

Patentes (4):

- AMPidentifier (Software, BR 512025005859-4, 2025) - Toolkit ML para predição de peptídeos antimicrobianos
- decryptAMP (Software, PC202600296, 2026) - Ferramenta bioinformática para descoberta de peptídeos encriptados
- Kit Detecção SARS-CoV-2 (BR1020210258438, 2021) - Oligonucleotídeos e método para detecção gênica
- Microscópio IoT (BR10202100250, 2021) - Tecnologia IoT para aquisição de imagem/vídeo

Prêmios & Honrarias:

- New Generation Bioinformatician Award, X-Meeting 2025 (Associação Brasileira de Bioinformática)
- Jovem Geneticista do Nordeste, Sociedade Brasileira de Genética (SBG) 2016
- Melhor Pôster, XIII Jornada PPG Genética & Biologia Molecular, UFPE 2023
- 2º Lugar (Genética Forense), XXI Encontro de Genética do Nordeste (ENGINE) 2016

Travel Grants:

- ICTP-SAIFR 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 (múltiplos anos)

Outros Reconhecimentos: Certificado de Excelência em Peer Review (Elsevier 2024) | Menção Honrosa em Genética Humana & Forense (Congresso Científico UFPE 2017) | Menção Honrosa Olimpíada de Química (Pernambuco 2013)

CERTIFICAÇÕES

AI/ML: Generative AI (MIT) | Oracle Cloud AI | Azure AI & ML (Microsoft) | Advanced Gemini (Google) | ML & Analytics (PUC-Rio) | Advanced Data Analysis (PUC-Rio)

Infraestrutura: 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)

IDIOMAS

Português (Nativo) | Inglês (Avançado)