Mariana Costa Almeida | Engenheira de IA Sênior | LLMs, MLOps e Visão Computacional

Rua dos Pinheiros, 1234, Apto 56, São Paulo, São Paulo, Brasil, 05422-001
+55 11 98765-4321 • mariana.almeida.ai@example.com • https://marianacosta-ai.dev
@marianacostaai • @marianacostaai • @marianacostaai

Informações Básicas

Engenheira de IA com mais de 7 anos de experiência no desenvolvimento e implantação de soluções de aprendizado de máquina em produção. Especializada em modelos de linguagem, visão computacional e pipelines de MLOps, com forte atuação em arquiteturas escaláveis e liderança técnica.

- Projetou e implantou sistemas de recomendação e busca semântica com LLMs, reduzindo o tempo de resposta em 40%.
- Experiência sólida em Python, PyTorch, TensorFlow, AWS e Kubernetes, com foco em reprodutibilidade e monitoramento contínuo.
- Liderou equipes multidisciplinares e programas de mentoria, promovendo boas práticas de IA responsável e experimentação ágil.

Formação

Universidade de São Paulo **https://www.usp.br**

Mestrado, Ciência da Computação com ênfase em Inteligência Artificial, Nota: 3.9/4.0 *mar. de 2018–dez. de 2020*

- Dissertação sobre modelos de linguagem aplicados à sumarização de documentos jurídicos.
- Publicação em simpósio nacional e colaboração com laboratório de IA da universidade.
- Monitoria em disciplina de aprendizado profundo por dois semestres.

Cursos: Aprendizado de Máquina Avançado, Processamento de Linguagem Natural, Visão Computacional, Aprendizado por Reforço, Otimização Convexa

Universidade Estadual de Campinas **https://www.unicamp.br**

Bacharelado, Engenharia de Computação, Nota: 3.7/4.0 *mar. de 2014–dez. de 2017*

- Formação com foco em sistemas distribuídos, computação em nuvem e fundamentos de IA.
- Trabalho de conclusão de curso sobre detecção de anomalias em séries temporais.
- Participação em grupo de estudos de aprendizado de máquina e maratonas de programação.

Cursos: Algoritmos e Estruturas de Dados, Sistemas Operacionais, Banco de Dados, Redes de Computadores, Estatística e Probabilidade, Inteligência Artificial

Experiência Profissional

Lumina AI Brasil **https://www.luminaai.com.br**

Engenheira de IA Sênior *jan. de 2022–Atual*

- Liderou a construção de uma plataforma de busca semântica e geração aumentada por recuperação (RAG) para o setor jurídico, atendendo mais de 50 mil usuários.
- Desenvolveu pipelines de ingestão, indexação e avaliação de LLMs com LangChain, FAISS e métricas de qualidade automatizadas.
- Implantou monitoramento de modelos em produção com Prometheus, Grafana e Evidently, reduzindo incidentes em 35%.
- Coordenou equipe de 6 engenheiros e cientistas de dados, definindo arquitetura, padrões de código e roadmap técnico.
- Promoveu práticas de IA responsável, incluindo revisão de vieses e documentação de modelos.

Palavras-chave: LLMs, RAG, MLOps, Liderança, Arquitetura

DataSemente Analytics **https://www.datasemente.com.br**

Engenheira de Machine Learning *fev. de 2020–dez. de 2021*

- Desenvolveu modelos de propensão de compra e churn para clientes de varejo, com ganho de 18% em precisão sobre a linha de base.
- Criou features e pipelines de treinamento em Spark e scikit-learn, processando mais de 2 TB de dados diariamente.
- Automatizou o deploy de modelos com Docker, Airflow e AWS SageMaker, reduzindo o tempo de entrega de 2 semanas para 3 dias.
- Colaborou com times de produto e negócio para traduzir requisitos em soluções de IA mensuráveis.

Palavras-chave: Machine Learning, Spark, SageMaker, Churn, Varejo

Inova Analytics **https://www.inovaanalytics.com.br**

Cientista de Dados Júnior *jan. de 2018–jan. de 2020*

- Construiu dashboards e modelos estatísticos para apoiar decisões de marketing e operações.
- Realizou análises exploratórias e testes A/B em campanhas digitais, identificando oportunidades de otimização.
- Desenvolveu scripts em Python e SQL para automatizar relatórios recorrentes e reduzir trabalho manual.
- Participou da migração de dados para ambiente cloud, garantindo qualidade e governança.

Palavras-chave: Python, SQL, A/B Testing, Estatística, Cloud

Idiomas

Português: Nativo ou Bilíngue — **Palavras-chave**: Nativo

Inglês: Proficiência Profissional Fluente — **Palavras-chave**: TOEFL iBT 112, Inglês técnico

Espanhol: Proficiência Profissional Limitada — **Palavras-chave**: Espanhol intermediário

Competências

Aprendizado de Máquina: Especialista — **Palavras-chave**: Python, PyTorch, TensorFlow, scikit-learn, XGBoost

Processamento de Linguagem Natural: Avançado — **Palavras-chave**: Transformers, LLMs, RAG, LangChain, Hugging Face

Visão Computacional: Avançado — **Palavras-chave**: OpenCV, YOLO, Detectron2, Segment Anything

MLOps e Cloud: Avançado — **Palavras-chave**: AWS SageMaker, Docker, Kubernetes, MLflow, Airflow

Engenharia de Dados: Intermediário — **Palavras-chave**: Spark, SQL, Kafka, dbt, Snowflake

Liderança Técnica: Avançado — **Palavras-chave**: Mentoria, Arquitetura de IA, Agile, IA responsável

Prêmios

Prêmio de Inovação em IA

Lumina AI Brasil *nov. de 2023*

Reconhecimento pelo desenvolvimento da plataforma de busca semântica com LLMs, que gerou economia estimada de R$ 2 milhões ao ano.

Melhor Artigo Científico

Simpósio Brasileiro de Inteligência Artificial *out. de 2021*

Premiação pelo artigo sobre aprendizado por reforço aplicado à otimização de recursos em nuvem.

Certificados

AWS Certified Machine Learning –Specialty **https://aws.amazon.com/certification/certified-machine-learning-specialty/**

AWS *mar. de 2022*

Google Cloud Professional Machine Learning Engineer **https://cloud.google.com/certification/machine-learning-engineer**

Google Cloud *jun. de 2023*

TensorFlow Developer Certificate **https://www.tensorflow.org/certificate**

TensorFlow *set. de 2021*

Publicações

Simpósio Brasileiro de Inteligência Artificial (SBIA) **https://sol.sbc.org.br/index.php/sbia**

Aprendizado por Reforço para Otimização de Recursos em Nuvem *out. de 2021*

- Propõe um agente de aprendizado por reforço para alocação dinâmica de recursos computacionais.
- Demonstra redução de custos operacionais em ambientes de nuvem com manutenção de SLA.
- Discute desafios de escalabilidade e reprodutibilidade em experimentos de longa duração.

Projetos

Assistente Jurídico com RAG **https://github.com/marianacostaai/a**

Assistente virtual que combina busca semântica e LLMs para responder dúvidas jurídicas com citações. *ja*

- Construiu pipeline de ingestão e chunking de documentos jurídicos com LangChain e FAISS.
- Implementou avaliação automatizada de respostas com métricas de fidelidade e cobertura.
- Desenvolveu API em FastAPI e interface web para demonstração do produto.

Palavras-chave: RAG, LLMs, LangChain, FAISS, FastAPI

Classificador de Imagens Médicas **https://github.com/marianacostaai/classifi**

Modelo de visão computacional para triagem de exames de imagem com foco em transparência. *m*

- Treinou e avaliou redes neurais convolucionais para classificação de exames de raios-X.
- Aplicou técnicas de aumento de dados e explicabilidade com Grad-CAM.
- Publicou resultados em workshop de IA aplicada à saúde.

Palavras-chave: Visão Computacional, PyTorch, Grad-CAM, Saúde, CNN

Pipeline de MLOps para Recomendação **https://github.com/marianacostaai/pipe**

Infraestrutura reprodutível para treinamento, versionamento e deploy de modelos de recomendação.

- Orquestrou treinamentos com Airflow e versionamento de dados e modelos com MLflow.
- Configurou deploy em Kubernetes com monitoramento de deriva de dados.
- Reduziu o tempo de treinamento e validação de modelos em 30%.

Palavras-chave: MLOps, Airflow, MLflow, Kubernetes, Recomendação

Interesses

Pesquisa em IA: LLMs, Aprendizado por Reforço, Ética em IA

Esportes: Corrida, Ciclismo, Natação

Música: Violão, Piano

Voluntariado

PyLadies São Paulo **https://www.pyladies.com**

Mentora Técnica *mar. de 2022–Atual*

- Mentora mulheres em transição de carreira para tecnologia, com foco em Python e fundamentos de IA.
- Organiza workshops de introdução a machine learning e boas práticas de projetos.
- Apoia a revisão de currículos e a preparação para entrevistas técnicas.