Yanka Ribeiro

Dresden, Germany | ynk9a@outlook.com | LinkedIn | GitHub | Personal Website

SKILLS & INTERESTS

Skills: MLOps; Cloud Computing (GCP); Scientific Research & Experimentation; Python; C++; SQL; TensorFlow; PyTorch; OpenCV; Scikit-learn; Hugging Face; SpaCy; Git; Docker; Seaborn.

Interests: **Machine Learning Engineering**: Computer Vision, Image Processing, Computer Graphics, and AI/ML Research.

WORK EXPERIENCE

HVAR Consulting **São Paulo, Brazil**

Machine Learning Engineer *May 2024 - Present*

- Developed and deployed advanced computer vision algorithms on the Google Cloud Platform (Vertex AI), achieving an F1-Score rate of 89% in image classification tasks across an imbalanced dataset of over 50000 images and videos, optimizing processing time by 40%.
- Developed a pipeline for automated SQL query generation based on user input leveraging Gemini model capabilities, training with extensive tabular data stored in BigQuery, achieving 99% accuracy in translating natural language to SQL queries and reducing query development time.
- Collaborated with cross-functional teams to deliver scalable ML solutions hosted on Cloud Run, improving operational efficiency for energy and safety compliance projects.

IBM Consulting **São Paulo, Brazil**

Data Scientist *Apr 2022 - Dec 2022*

- Collaborated directly with clients to understand business requirements, translate them into technical solutions, and deliver actionable insights, ensuring alignment with goals.
- Developed semi-supervised learning models for multi-class classification, using NLP text manipulation to analyze expense patterns, achieving an average precision of 90% in identifying the most common budget categories.
- Leveraged Python libraries like Pandas for data manipulation, Scikit-learn for semi-supervised learning models, and NLTK for NLP tasks.
- Built and deployed the entire pipeline on the client's on-premises infrastructure, utilizing Docker for containerization maintaining data security and compliance.

EDGE Research Center **Maceió, Brazil**

Machine Learning Engineer *Jan 2021 - Feb 2022*

- Designed and implemented a rule-based NLP engine using spaCy to automatically extract key specifications from government bid documents (often exceeding 50 pages), enabling the development of a product recommendation system that matched client requirements with the company's product catalog. This solution reduced bid losses by 65%.
- Fine-tuned BERTimbau (PT-BR) using domain-specific vocabulary expansion to enhance complex feature extraction, improving the accuracy of matching technical specifications for CT equipment and adapting to dynamic tender requirements.
- Partnered with backend developers to build robust web solutions, leveraging RESTful APIs and microservices architecture to deploy models and successfully delivering web applications hosted on the client's domain.

EDUCATION

Technical University of Dresden **Dresden, Germany**

MS. Computational Modelling and Simulation (track Visual Computing) *Graduation Date: Oct 2026*

Federal University of Alagoas **Maceió, Brazil**

BS. Computer Science *Graduation Date: Mar 2024*

Thesis title: Evaluation of deep metric learning methods for the diagnosis of human visceral leishmaniasis.

CERTIFICATES

Google: Certified as *Professional Machine Learning Engineer* (2025)

Certified as *Associate Machine Learning Specialist* (2022)

IBM: Certified as *Associate Data Scientist*

RESEARCH EXPERIENCE

Google Explore CSR **Remote**

LATAM Research Mentoring Program *Feb 2024 - Jul 2024*

- Evaluated Vision Language Models (VLMs) to describe and annotate artifacts and metadata in X-ray datasets, enabling future bias analysis in computer-aided diagnosis.
- Implemented the project using computer vision, NLP, and pre-trained multimodal healthcare models in Python with PyTorch.
- Presented a 10-minute summary of the first 6 months of research progress. Advisor: Phd Enzo Ferrante (CONICET)

LaCCAN Research Laboratory **Maceió, Brazil**

Research Assistancy *Aug 2023 - Sep 2024*

- Participated in creating a systematic review on AI algorithms for evaluating auditory brainstem response in time and frequency domains to facilitate early diagnosis of auditory alterations.
- Presented the research as a poster at a speech therapy conference. Prepared the review for publication, currently under submission. Advisor: Phd Raquel Cabral (UFAL)
- Built and evaluated four deep metric learning methods to accurately diagnose human visceral leishmaniasis using blood smear images, achieving over 95% in key performance metrics (recall, precision, and F1-score).
- Refined the work and published it in a national symposium on applied computing in healthcare.
- Utilized computer vision, image processing, and classification techniques, implemented in Python with PyTorch, alongside CNNs, SVM, and PCA for feature analysis. PhD Fabiane Queiroz (UFAL)

UNIFESP Medicine Faculty **São Paulo, Brazil**

Research Project in Computer Vision/Ophthalmology *Sep 2021 - Aug 2022*

- Funded by FAPESP, the research aims to enhance ocular toxoplasmosis diagnosis and monitoring, advancing public health in Brazil. Advisor: Phd Rubens Belfort (UNIFESP), Phd Luis Nakayama (MIT)
- Developed a image segmentation algorithm to track ocular toxoplasmosis using the BR-OPHTSET database, comparing neural networks like ResNet, VGG, DenseNet, and EfficientNet with various optimizers.
- Achieved 95% IOU and produced a research article to present at the conclusion of the project.