

Bella Nicholson

A resourceful and results-oriented Machine Learning Engineer (3+ years) with a proven track record of building, optimizing, and deploying production-ready ML solutions across diverse tech environments. A life-long learner who's well-versed in LLM inference and edge deployment. Excited about the intersection between software engineering and AI research.

EXPERIENCE

Machine Learning Engineer // Brenntag //  Nov 2023 – Present

 Amsterdam, Netherlands

At [the world's leading chemical distributor](#), I deployed and maintained ML products across [70+ countries](#). Thus far, I've:

- Migrated a €30M+ annual revenue AI assistant to a **more cost-effective and secure AWS** platform, reducing operational costs
- Developed a **real-time notification system** to monitor critical ML jobs and model metrics, **improving system visibility** and reliability
- Standardized quality controls across 15+ ML project components through the creation of a CLI package prototype

Machine Learning Consultant // Deloitte //  Sept 2021 – Oct 2023

 Amsterdam, Netherlands

As a contracted ML engineer, I optimized and implemented ML solutions for diverse clients, delivering:

- Increased system robustness of a **Dutch e-classified ads platform's "For You" recommendation engine** with ML Engineering best practices
- Launched a self-paced, ML-focused coding training website to standardize and improve code quality across [Deloitte NL](#)
- Centralized model storage and experiment tracking across environments for a [German steel conglomerate](#), improving model reproducibility

Machine Learning Research Intern // Crunchr //  Jan 2020 – Aug 2020

 Amsterdam, Netherlands

At [a people analytics platform](#), I conducted graph-based representation learning research for the **development of ML products**:

- Built a proof of concept **representation learning** process to encode relational database entities for **improved downstream ML performance**
- Demonstrated approach validity by applying deep neural networks to downstream classification tasks on process outputs

Computer Vision Intern // Cubelizer //  June 2017 – July 2017

 Madrid, Spain

As part of [a Google-backed edge computer vision startup](#), I improved customer detection by 12% for retail space optimization:

- Developed a video stream-based **object detection** method in compliance with EU privacy regulations
- Applied image processing and classical machine learning techniques to low-resolution images

PROJECTS

Pocket Multi-Modal Large Language Model


- Deployed a **custom embedded, vision-text foundation model** across various iOS devices (laptop, phone, tablet)
- Extended [an open source hardware-optimization framework](#) to quantize and optimize [a new multi-modal Large Language Model](#)
- Documented the project implementation, including application solution prototyping, in [a detailed 4-part blog post series](#)

Transformers Decoded: A Guide to Optimizing Large Language Models

- Developed [a comprehensive study guide on Large Language Models \(LLMs\)](#), explaining underlying concepts and modern optimization strategies
- Covered LLM inference **optimization techniques** (**speculative decoding, flash attention, continuous batching; etc.**) for efficient deployment

EDUCATION

Master of Science, Artificial Intelligence

 [University of Amsterdam](#)

 Cum laude (8.0/10.0)


 Sept 2018 – Dec 2020


 Amsterdam, Netherlands

- Courses on AI, including Deep Learning, Computer Vision, Natural Language Processing, Information Retrieval, and Reinforcement Learning
- Thesis on ["Interpretable Representation Learning for Relational Data"](#) in collaboration with Crunchr

Bachelor of Science, Biomedical Engineering

 [The College of New Jersey](#)

 Magna cum laude (3.8/4.0)

 Sept 2014 – May 2018

 Ewing, New Jersey, USA

SKILLS

Programming Languages & Tooling

Python, Git (2016-present), Bash (2018-present), Terraform (2023-present), SQL (2019-present), Docker (2022-present)

MLOps Platforms

Amazon Web Services (2021-present), Google Cloud Platform (2022-present), Databricks (2022-present)

ML Frameworks

PyTorch (2018-present), PySpark (2022-present), FastAPI (2021-present), Tensorflow (2021-2022), Keras (2022)