

Milla Nielsen

Berkeley, CA | millaenielsen@gmail.com | 510-725-8858 | [linkedin.com/in/milla-nielsen](https://www.linkedin.com/in/milla-nielsen)
github.com/millaenielsen

Education

University of California, Los Angeles, BS in Statistics and Data Science Sept 2021 – June 2025
Minor in Data Science Engineering (GPA: 3.86/4.0)

- **Coursework:** Algorithms, Data Mining, Applied Advanced Statistics, Computational Statistics, Probability Theory, Linear Algebra, Multivariable Calculus, Machine Learning in Python, Linear Models, Generative AI, Markov Chain Monte Carlo, Deep Learning
- **Awards:** UCLA ASA DataFest 2024 Finalist, UCLA ASA DataFest 2023 Best Use of External Data (Honorable Mention)

Experience

Computer Vision Data Science Intern, Prenatal (prenatal AI start-up) – July 2025 – September 2025
Copenhagen, Denmark

- Developed preprocessing components for medical image pipeline, achieving **95%+ accuracy** in artifact removal and improving model reliability
- Applied computer vision and OCR methods, using statistical testing to select optimal algorithms for large-scale deployment
- Delivered **production code** for deep learning pipeline with **Git** version control for research team

Research Intern, Pioneer Centre for AI – University of Copenhagen July 2024 – Aug 2024

- Developed MRI-based machine learning pipeline for Alzheimer's detection, achieving **0.84 AUC** to enable earlier diagnosis
- Engineered biomarker features with **3D CNN (PyTorch, transfer learning)** and optimized XGBoost via cross-validation + hyperparameter tuning
- Assessed performance with ROC analysis and statistical validation, contributing findings to conference paper

Communications Data Science Lab Assistant, UCLA – Los Angeles, CA Aug 2023 – June 2024

- Enhanced Russian speech-to-text via **OpenAI Whisper** fine-tuning for media bias research applications
- Benchmarked Whisper variants using BLEU score + WER, visualizing results with **R (ggplot2)** and **Python (Matplotlib)**
- Scaled **GPU-accelerated** pipeline on UCLA's Linux cluster, transcribing 1,000+ multi-hour audio files

Technologies

Languages: C++, Python, R, SQL

Cloud/Tools: GCP (Cloud Storage, BigQuery), Git/GitHub, Tableau, Jupyter, RStudio, Linux, LaTeX, Unix Bash

ML/Data: PyTorch, TensorFlow, NumPy, Pandas, Scikit-learn, Hugging Face, OpenCV, ggplot2, Matplotlib

Leadership

Co-President, National Student Data Corps - UCLA Feb 2023 - June 2025

- Scaled UCLA's NSDC chapter from 200 → 900+ members through 50+ workshops, hackathons, and training programs providing hands-on experience in Python, Tableau, NLP, and ML

Head of Consulting, UCLA Data Resolutions March 2024 - March 2025

- Managed 5+ consulting teams (35+ analysts total) serving 15 startup clients, building scalable processes to deliver data-driven solutions across quarterly project cycles

Publications

Milla E. Nielsen, Mads Nielsen, Mostafa Mehdipour Ghazi, "Assessing the efficacy of classical and deep neuroimaging biomarkers in early Alzheimer's disease diagnosis," Proc. SPIE 13407, Medical Imaging 2025: Computer-Aided Diagnosis, 1340722 (4 April 2025); <https://doi.org/10.1117/12.3046952> April 2024