

KATARZYNA SZYMANIAK

Machine Learning Research Scientist | Health AI, Wearables and Biosignals

📍 Edinburgh, UK 🌐 github.com/KatarzynaSzymaniak 🏠 katarzynaszymaniak.github.io
✉ k.szymaniak.ai@gmail.com 🔗 linkedin.com/in/katarzyna-szymaniak-ai 🎓 Google Scholar

SUMMARY

Machine learning researcher (PhD in Biomedical AI, Edinburgh, 2026) building biosignal and wearable-sensing models that stay reliable across users, sessions and real-world distribution shift. **Wearables industry research:** Research Scientist Intern at Meta Reality Labs (CTRL-labs), working on multimodal sensor fusion (sEMG + IMU) and few-shot personalisation for continuous navigation in neural interfaces. **Robustness for wearables:** released an open EMG benchmark isolating arm-position shift (*Nature's Scientific Data*) and developed domain generalisation methods for calibration-free control (EurIPS 2025). **Data-efficient adaptation:** human-in-the-loop active learning for long-term adaptation of myoelectric decoders using uncertainty-based query strategies. **End-to-end:** from human-subject protocol design and data collection through modelling and evaluation. **Now:** whether **health-audio foundation models** rely on **physiology** or on **device and site confounds** (**counterfactual evaluation, RL post-training**), in collaboration with NVIDIA Research.

RESEARCH EXPERIENCE

Independent Research, in collaboration with NVIDIA Research

Machine Learning Researcher, Source Shortcuts in Health-Audio Language Models

Edinburgh, UK

Aug 2026 – Present

- Investigating whether **health-audio LLMs** diagnose from physiology or from **recording-device and site confounds**, extending my thesis on confounding factors from EMG to clinical audio (respiratory and auscultation sounds).
- Designing **counterfactual evaluation, encoder probing and cross-dataset generalisation** tests to expose shortcut learning in multimodal foundation models; planned next steps are device-invariant self-supervised pretraining, RL post-training (GRPO, DPO) and agentic health-audio evaluation.

Meta Reality Labs (CTRL-labs)

Research Scientist Intern, Continuous Navigation (Host: Fabio Stefanini)

New York, NY

Sep 2022 – Mar 2023

- Extended the team's wrist-worn sEMG decoding stack to **multimodal sEMG + IMU sensor fusion** for neuromotor AR/VR input: built inertial signal preprocessing, trained **early-fusion decoders**, and diagnosed **modality imbalance** between inputs.
- Built an **end-to-end personalisation pipeline** for new wrist gestures, from designing the human-subject data-collection protocol to **few-shot per-user calibration** of a pretrained Conformer-based sEMG decoder.
- Evaluated **cross-session generalisation** under band re-donning with per-user **learning curves** over calibration budgets; presented results and limitations at the team's technical deep dive.
- Context: the team's cross-user sEMG decoding work was published in Nature (2025) and underpins the Meta Neural Band.

University of Edinburgh, School of Informatics

Doctoral Researcher, Biomedical AI

Edinburgh, UK

Sep 2020 – Jan 2026

- Led thesis research on **robust myoelectric control** for upper-limb prostheses: why EMG decoders fail outside the lab (limb position effect, long-term signal drift) and how to mitigate these confounding factors with **domain generalisation, active learning and adaptive probabilistic modelling**.
- Ran multi-session human-subject studies, including amputee participants, working across the full stack from sensor hardware and anatomy to signal processing and ML modelling.
- Built reusable data-processing and evaluation pipelines for EMG studies, **adopted across the lab**.

SELECTED PUBLICATIONS

* equal contribution

Position Domain Generalisation in Myoelectric Control

K. Szymaniak, K. Nazarpour, H. Gouk

EurIPS 2025 Workshop | extended version under review at TMLR

- Proposed **DR-cGMM**, a domain-regularised class-conditional Gaussian mixture model that generalises EMG gesture decoding to unseen arm positions with zero data from those positions, via a covariance regulariser that shares structure across positions.
- Outperformed ERM, MMD and CORAL, **the only method to do so**, showing that principled statistical modelling beats deep learning in the **low-data, per-user regime typical of wearables**.
- Designed a **leakage-free benchmark**: leave-one-domain-out model selection with nested cross-validation and Friedman/Nemenyi significance testing over **about 40,000 trained models**.

EMG Dataset for Gesture Recognition with Arm Translation (GREAT)

I. Kyranou*, K. Szymaniak*, K. Nazarpour | open dataset, 21,000+ accesses

Scientific Data (Nature Portfolio), 2025

- Designed the experiment and acquisition protocol for an open sEMG + hand-kinematics benchmark (**16-channel sEMG at 2 kHz, 18-sensor CyberGlove II**) recorded over a 3×3 grid of 9 arm positions across vertical and horizontal planes.
- Quantified the shift: **accuracy dropped by up to 10%** across positions, and position itself was **decodable from the signal**.

(about 80%); showed that position-aware hierarchical decoding fails outside an i.i.d. setting (position encoder 58%), motivating the domain generalisation work above.

Recalibration of Myoelectric Control with Active Learning

K. Szymaniak, A. Krasoulis, K. Nazarpour

Frontiers in Neurorobotics, 2022

- **Introduced active learning to myoelectric control for the first time:** the model queries labels only for windows ranked most informative (least-confidence, margin, entropy).
- Achieved **86% accuracy from 3.2 minutes** of actively queried data, versus 82% for both a full calibration session and random sampling (12 able-bodied participants; about 2 points gain in 2 transradial amputees; offline simulation).
- Identified a **failure mode of diversity-based batch selection:** it degraded margin sampling, likely because the most informative boundary samples cluster and are discarded as redundant.

Deep Feature Learning from Electromyographic Signals for Gesture Recognition Systems

W. Zhong, X. Jiang, K. Szymaniak, M. Jabbari, C. Ma, K. Nazarpour

IEEE TNSRE, 2026 (survey)

EARLIER RESEARCH

- **Predicting behaviour from brain data** (University of Edinburgh, 2021; Prof. Oisín Mac Aodha, with the Centre for Discovery Brain Sciences): decoded the location of freely behaving mice from in vivo deep-brain calcium imaging, using class-activation-style attribution to identify the neuronal populations driving predictions.
- **Sequence models for EEG** (Swansea University, 2020; Dr Jingjing Deng, Prof. Xianghua Xie): built deep sequential models for EEG in decision deadlocks, handling temporal dependencies, non-stationarity, cross-channel connectivity and noise, towards brain-computer interfaces.
- **HoloLens neuroimaging for medical education** (Swansea University, 2019; Best Student Project award): rendered MRI brain scans as 3D meshes and deployed them on Microsoft HoloLens via Unity as a mixed-reality teaching tool.

EDUCATION

PhD, Biomedical Artificial Intelligence, University of Edinburgh, UK

Sep 2020 – Jan 2026

Thesis: Mitigating Confounding Factors in Myoelectric Control Through Adaptive Modelling and Learning. Supervisors: Prof. Kianoush Nazarpour, Prof. Timothy Hospedales.

MRes, Biomedical Artificial Intelligence, University of Edinburgh, UK

Sep 2019 – Sep 2020

MRes, Visual Computing, Swansea University, UK

Sep 2018 – Sep 2020

BSc (Hons) Computer Science, First Class Honours, Swansea University, UK

Sep 2015 – Jul 2018

Oxford Machine Learning Summer School (OxML), University of Oxford, UK

Aug 2022

TECHNICAL SKILLS

Biosignals and Sensors: sEMG, IMU, EEG, MRI, health audio; wearable sensing, Delsys Trigno, CyberGlove II hand kinematics; myoelectric control, gesture recognition, neural interfaces

Human-Subject Research: protocol design, multi-session data collection, amputee participants, open dataset release

Signal Processing: biomedical signal processing, spectral and time-domain features, artefact handling, real-time windowing

Machine Learning: time-series deep learning (CNN/TCN, RNN/LSTM, Transformers, Conformers), time-series classification, multimodal sensor fusion, self-supervised and representation learning, transfer learning, domain adaptation and generalisation, OOD generalisation, personalisation, few-shot and active learning, uncertainty estimation, interpretability

LLMs and Post-Training: SFT, GRPO, PEFT/LoRA, LLM and agent evaluation, counterfactual evaluation

Statistics and Evaluation: experimental design, benchmark design, model selection, hyperparameter tuning, cross-validation, significance testing, cross-user and cross-session evaluation

Programming and MLOps: Python, PyTorch, PyTorch Lightning, Hydra, scikit-learn, NumPy, SciPy, Pandas, SQL, MATLAB, Java, TensorFlow; MLflow, Weights & Biases, Docker, Kubernetes, Git

TEACHING

Teaching Assistant, Machine Learning Practical (MLP), University of Edinburgh

Sep 2024 – Jun 2025

Teaching Assistant, Entrepreneurship and Innovation, University of Edinburgh

Sep 2023 – Jul 2024

Mentored student teams building ML-driven MVPs for real-world health and technology problems.

Lab Demonstrator, Java OOP, Data Structures and Algorithms, Swansea University

Jan 2018 – Sep 2019

SERVICE AND LEADERSHIP

- Peer Reviewer, IEEE TNSRE; Frontiers in Neurorobotics Nov 2024 – Aug 2025
- Co-organiser, Women in Machine Learning (WiML) Workshop at NeurIPS 2021 Dec 2021
- Co-organiser, two neuroprosthetics co-creation and co-design workshops, University of Edinburgh 2021, 2022
- Research collaborations: Centre for Discovery Brain Sciences, Edinburgh (2020); CUBRIC, Cardiff (2018–2020)