Sepehr Behtaji *ML Engineer | Computer Vision & Applied AI (Neurodiagnostic + FinTech)*

Sepehr.Behtaji@gmail.com
07535658759
UK
linkedin
github
Google Scholar

Profile

Machine Learning / AI Engineer focused on computer vision and applied AI, with experience delivering end-to-end ML systems from data preparation and modelling through evaluation, explainability, and deployment. Strong Python background (feature engineering, supervised learning, validation, and model interpretation). Combines hands-on ML engineering with technical product thinking—translating user and business needs into reliable, measurable AI features.

Education

2023/01 – 2024/05
Middlesbrough, UK

MSc in Bioinformatics; Data science applied to biology (Distinction)
Teesside University
Thesis: *Predictive modelling of high-dimensional biological datasets*
- Integrated multiple large-scale sequencing datasets to analyse relationships between regulatory features and system-level output metrics
- Built, evaluated and interpreted machine-learning models (XGBoost, Random Forest) in Python to identify key predictive features
- Applied model interpretability techniques (SHAP) to translate complex models into clear, explainable insights.

2019/09 – 2020/12
Nancy, France

MSc in RNA and Enzyme Sciences (Merit)
University of Lorraine
Oxytocin-related lncRNAs and target genes in breast cancer (expression profiling)
- Bioinformatics: gene selection from public datasets + pathway enrichment (KEGG/Enrichr)
- Wet lab: RT-qPCR workflow (RNA extraction, cDNA synthesis, primer design, melt-curve/gel QC)

2014/07 – 2018/07
Tonekabon, Iran

BSc in Cellular and Molecular Biology, Genetics (First Class)
Azad University of Tonekabon

Experience

2025/05 – present
London, United Kingdom

AI / ML Engineer & Technical Product Manager
GoodFolio LTD
- Developed and internally validated high-accuracy computer-vision models for GoodMind's CT/MRI neurodiagnostic platform across stroke, dementia and brain tumour use cases.
- Built CNN, 3D medical-imaging and classifier workflows for disease classification, abnormality localisation and diagnostic-support outputs.
- Trained and evaluated dementia MRI models across multiple datasets, using confusion matrices, cross-dataset testing and Grad-CAM heatmaps to assess performance and explain model decisions.
- Developed CT-stroke classification workflows to distinguish normal scans, ischemic stroke and haemorrhagic/bleeding cases, supporting urgent neuroimaging triage use cases.
- Created ML-ready imaging cohorts from DICOM scans, clinical labels, metadata, timestamps and outcome proxies to support model training, evaluation and validation.
- Translated clinical needs, dataset requirements and AI capabilities into product features, validation plans, roadmap priorities and pilot-ready workflows.
- Developed and deployed AI agents for automated client outreach using Google ADK, Gemini, Vertex AI, Model Garden, Cloud Run and Vertex AI Agent Engine.
- Built and evaluated ML components for rule-to-check generation and layout/vision-aware checks, producing explainable, auditable inspection outputs for regulated fintech/regtech firms.
- Collaborated across AI, product, security and operations teams, while also supporting ISO 27001 audit-readiness, risk tracking and compliance documentation.

2024/01 – 2024/06
Darlington, UK

MSc Researcher – Data Modelling & Analytics (MSc dissertation project)
Teesside University

- Built and evaluated supervised ML models on large, complex datasets to identify predictive and explainable patterns in gene-level translation outcomes across two treatment conditions.
- Led an independent analytical project integrating multiple high-dimensional datasets to study relationships between regulatory features and system-level outputs
- Designed and executed data processing pipelines for large sequencing-derived datasets, focusing on data quality, alignment and aggregation
- Built and compared interpretable machine-learning models (Random Forest, XGBoost, AdaBoost) to identify key predictive features
- Applied SHAP-based interpretability to translate complex models into clear, explainable insights

2023/03 – 2025/04
London, UK

Contractor (part-time) – Data & Machine Learning Engineer
ATLASI Ltd (*Client delivery / services*)

- Built and maintained ETL pipelines for high-volume datasets supporting client analytics, reporting, and model evaluation.
- Used SQL and Python for ad-hoc analysis, data validation, and exception reporting.
- Implemented data quality checks to reduce defects, improve cross-system consistency, and strengthen traceability of key fields.
- Supported release/testing cycles by verifying data integrity across environments and preventing reporting/pipeline regressions.
- Wrote and maintained end-to-end Gherkin test packs to validate workflows and pipeline behaviour before release.
- Worked in cloud environments to run scalable, repeatable processing and scheduled validation tasks.

2021/10 – 2022/12
Tehran, Iran

Data Scientist / Bioinformatics Researcher
Geniran Research Laboratory

- Built machine-learning models on large, complex datasets to identify predictive and explanatory patterns
- Curated and integrated structured and unstructured features, developing experience in feature selection and model interpretation
- Collaborated with multidisciplinary teams to translate analytical findings into testable hypotheses and insights

2020/04 – 2020/12
Nancy, France

Research Intern – Data Analysis & Validation
University of Lorraine

- Combined computational analysis with experimental validation, strengthening skills in data quality control and interpretation.

Professional Certificates

Python programming
Sematec IT Training Center

SQL Server 2022 Database Implementation
Sematec IT Training Center

Google Agent Development Kit (ADK)-Build & Deploy AI Agents
Udemy

Publications

2026/05/10

High-Accuracy Lung Region Segmentation and Injury Classification Using Deep Networks
Scientific Reports

2026/04/20

The adaptive engagement framework: enhancing banking customer experience through AI-powered invisible marketing
Scientific Reports

2021/03/19 **Identification of oxytocin-related lncRNAs and assessment of their expression in breast cancer**
Scientific Reports

Core Skills

Machine Learning & AI
- Supervised learning: classification, regression, model evaluation and validation
- Algorithms: XGBoost, Random Forest, AdaBoost, SVM, KNN, Gradient Boosting, Logistic Regression
- Feature engineering, feature selection, dimensionality reduction and correlation analysis
- Model interpretability and explainability: SHAP, feature importance, bias assessment
- Cross-validation and performance reporting: ROC-AUC, F1-score, precision, recall, Dice coefficient
- Class imbalance handling and sampling strategies, including Random OverSampler

Medical Imaging & Computer Vision
- CT/MRI medical-imaging analysis for stroke, dementia and brain tumour use cases
- Explainable medical-imaging AI using Grad-CAM heatmaps and confusion-matrix analysis
- CT/MRI classification workflows for dementia staging, stroke subtype detection and tumour-support use cases
- 3D medical-imaging workflows, CNN-based classification and diagnostic-support modelling
- Image segmentation and multi-class classification using U-Net and CNN backbones
- CNN architectures: ResNet50, VGG16, Xception
- Hybrid modelling: CNN feature extraction with classical ML classifiers
- Imaging-derived feature analysis, abnormality localisation and model benchmarking

AI Product, Validation & Stakeholder Communication
- Translating clinical, technical and business requirements into AI product features
- Technical product management, roadmap support and pilot-ready workflow design
- Documentation of data structures, model assumptions, rules and processes
- Communicating complex analytical results to technical and non-technical stakeholders
- Statistical reasoning, experimental design and technical reporting

Programming & Tools
- Python and R for data processing, automation and machine-learning workflows
- Python libraries: pandas, NumPy, scikit-learn, PyTorch
- Data visualisation and reporting: Matplotlib, Seaborn
- Version control: Git
- Collaboration and delivery tools: Jira
- Advanced Excel for data validation, manipulation and reporting

ML Data Infrastructure & Cloud
- Data extraction, cleansing, preprocessing and validation for large real-world datasets
- Building and optimising data pipelines for high-dimensional and multimodal datasets
- Handling structured, unstructured and imaging-derived data
- SQL database querying, reconciliation and troubleshooting
- Cloud and scalable processing environments: Google Cloud, AWS, Azure, HPC
- Workflow automation and reproducibility using Nextflow and Jupyter Notebook

AI Agents & Product Automation
- AI agent development, deployment and workflow orchestration
- Google ADK, Vertex AI Agent Engine, Gemini on Vertex AI, Model Garden and Cloud Run
- Agentic outreach automation and growth workflow support
- Production-oriented AI workflow design and cloud deployment