Roshani Nitin Pawar

+44 7586 475467 | roshani.n.pawar@gmail.com
LinkedIn | GitHub | Portfolio

Education

MSc Artificial Intelligence Jan. 2024 – May 2025
University of Aberdeen *Distinction*

- Coursework: Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, Research Methods
- Dissertation: "Vocabulary Through Time: Identifying Methodological Patterns in Temporal Distribution Analysis" (Supervisor: Dr Wei Zhao)

AI for Medicine Specialization (in progress) Jan. 2026 – Present
Coursera / deeplearning.ai (Stanford) *Online*

- Medical imaging analysis (chest X-rays, dermatology), clinical risk prediction models, treatment effect estimation

Master of Computer Applications July 2021 – Sept. 2023
Savitribai Phule Pune University, India *CGPA 8.5/10.0*

Bachelor of Computer Applications July 2018 – July 2021
Savitribai Phule Pune University, India *CGPA 9.6/10.0*

Research Experience

MSc Dissertation in Natural Language Processing Jan. 2025 – May 2025
University of Aberdeen *Supervisor: Dr Wei Zhao*

- Analyzed how LLM tokenizers process text from different time periods and sources, examining 6 tokenizers (GPT-2, GPT-3.5, GPT-4, Claude variants) on historical corpora spanning 1850s-2020s
- Applied statistical modeling and linear programming to quantify compression efficiency variations, finding 15-40% degradation on historical and domain-specific text compared to contemporary sources
- Relevant to clinical NLP: medical terminology evolves and reporting styles vary across institutions and time periods - tokenizer biases could affect systems processing diverse hospital documentation or longitudinal health records

Professional Experience

Software Engineer May 2025 – July 2025
University of Aberdeen, Medical Research (Reproductive Medicine) *Aberdeen, UK*

- Migrated OPIS IVF success rate calculator from Flask to Next.js, working with reproductive medicine researchers on a clinical decision support tool used globally
- OPIS predicts fertility treatment outcomes using HFEA registry data from 100,000+ cycles, incorporating patient age, ovarian reserve markers (AMH levels, antral follicle count), fertility diagnosis, and treatment protocol
- Built interface for clinicians to input parameters and view personalized success rate predictions - learned how presenting uncertainty (point estimates versus confidence ranges) affects patient counseling
- Discussions with researchers highlighted challenges in medical prediction: balancing statistical accuracy with clinical interpretability, and communicating limitations to patients making treatment decisions

Placement Programmer Jun. 2024 – Aug. 2024
Centre for Healthcare Randomised Trials (CHaRT), University of Aberdeen *Aberdeen, UK*

- Built automated messaging system for RCT participant communications in clinical trials research unit
- Worked with trial coordinators who explained RCT methodology: how randomization and blinding prevent bias, why intention-to-treat analysis maintains validity despite dropouts, and requirements for documenting protocol deviations and serious adverse events
- Developed Python scheduling system integrated with trial databases for appointment reminders and protocol-specific communications, reducing manual coordinator workload 40%
- PostgreSQL database with encryption and audit trails meeting NHS Scotland governance requirements, MHRA inspection standards, and Research Ethics Committee protocols

Senior Systems Associate Dec. 2021 – Dec. 2023
Infosys *Remote*

- Led agile teams (6-8 members) building enterprise data systems on Microsoft Azure platforms
- Developed data pipelines using Python and .NET with Azure ML for supply chain predictive analytics

Technical Projects

Medical Imaging Analysis (AI for Medicine Coursework) Sept. 2024 – Present

- Hands-on experience with medical imaging challenges: severe class imbalance in disease datasets, understanding why sensitivity-specificity tradeoffs differ for screening (prioritize sensitivity) versus diagnosis (balance both), and why model interpretability matters for physician trust
- Completed projects in chest X-ray classification and dermatology image analysis using TensorFlow and PyTorch, working with datasets where positive cases represent less than 5% of samples

Engagement with Biomedical AI Literature

Clinical NLP for radiology: Read Alex et al. (2019) "Text Mining Brain Imaging Reports" on EdIE-R's rule-based brain phenotype extraction from radiology text, Casey et al. (2021) "Natural Language Processing Applied to Radiology Reports" systematic review of NLP approaches across imaging modalities, and Casey et al. (2023) "Understanding the Performance and Reliability of NLP Tools" comparing neural versus rule-based methods for stroke phenotype extraction - found performance varies significantly across datasets and patient demographics

Healthcare AI fairness: Read Davidson et al. (2021) "Reporting Quality of Natural Language Processing Studies" on reproducibility challenges in clinical NLP and papers examining algorithmic bias in clinical decision support, particularly how performance gaps across demographic groups create equity concerns

Temporal bias in medical AI: Examined literature on how language model performance varies across time periods and how this relates to fairness concerns when processing historical medical records or documentation from different healthcare systems

Technical Skills

Programming: Python (NumPy, Pandas, Scikit-learn, SciPy, Matplotlib), SQL, Git

Machine Learning: TensorFlow, PyTorch, OpenCV, statistical modeling, experimental design

Development: FastAPI, Docker, Azure, PostgreSQL, RESTful APIs

Research: Literature review, hypothesis testing, data visualization, reproducible workflows

Biomedical Concepts from Professional Experience

Clinical trials (CHaRT): Worked with coordinators managing randomization and blinding procedures, learned why intention-to-treat analysis preserves trial validity, understood requirements for GCP compliance, serious adverse event documentation, and Research Ethics Committee protocols

Reproductive medicine (OPIS): Learned how IVF/ICSI success depends on ovarian reserve markers (AMH levels, antral follicle count), stimulation protocols, and patient age - discussions with researchers on challenges in communicating prediction uncertainty during patient counseling

Healthcare data governance: Experience with NHS Scotland data standards, GDPR requirements in research contexts, MHRA inspection protocols, participant consent management in clinical trials

Community Involvement

Technical Volunteer May 2020 – Present

Unique Educational Foundation, India *Remote*

- Built data collection and analysis tools for program evaluation at educational NGO
- Developed predictive models to identify students at risk of dropout for early intervention support

Interests

Swimming, taking pictures of things I find interesting, growing houseplants, and exploring the Scottish Highlands. Curious about how things work — from cameras to language models.