Sean Farrell | Durham, United Kingdom

Email | LinkedIn | website
sean.farrell2@durham.ac.uk | seanoliverfarrell | seanfarr788.github.io

A highly driven PhD candidate specialising in natural language processing (NLP) and big data analytics, grounded in a robust foundation of biomedical sciences. My academic and professional journey is marked by a proven record of impactful, peer-reviewed publications, successful interdisciplinary collaborations across international teams, and hands-on experience teaching NLP to diverse audiences. Driven by a passion for addressing critical public health challenges, my work combines computational innovation, technical expertise, and strong communication skills to transform data into actionable insights with a meaningful public health impact.

EDUCATION

PhD Computer Science (Natural Language Processing)
October 2021 – Present
Durham University
Biotechnology and Biological Sciences Research Council (BBSRC) funding to explore Natural Language Processing and deep learning methods to analyse over 10 million first opinion veterinary electronic health records (EHRs) from across the UK to understand critical public health matters. Key achievements include:

- **PetBERT Development:** Designed a novel domain-specific transformer language model.
- **Public Health Research:** Conducted studies on critical issues such as disease outbreak detection, socioeconomic disparities in premature mortality, and patterns of antimicrobial usage in alignment with antimicrobial stewardship guidelines.
- **Open Benchmark Dataset:** Established the first benchmark dataset for veterinary free-text EHRs, setting a new standard for research in this field.
- **Multimodal and Explainable AI:** Integrated free text and structured data with explainability tools to provide actionable insights into model decision-making processes.
- **Multinational Collaboration:** Coordinated a multinational study within the EU ENOVAT project, identifying barriers for antimicrobial stewardship guideline adoption.
- **International Protocol Leadership:** Leading the creation of global protocols for anonymisation, sharing of veterinary free-text EHRs and language models to promote ethical data use and collaboration.

Impact: My research bridges computational innovation and public health applications, advocates for open science principles and setting new standards for veterinary data and model sharing

Thesis: *Natural Language Processing for Early Detection and Mitigation of Critical Public Health Threats*

2:1 BSc (Hons) Biomedical Sciences
September 2018 – June 2021
University of Kent
Proficient in applying interdisciplinary knowledge across biology and healthcare to analyse and address health-related challenges. Experience in laboratory methods in Genetics, Microbiology, Biochemistry, and Immunology

Thesis: *Antimicrobial usage in hospitalised SARS-CoV-2 patients and the impact on the Gut Microbiome*

PUBLICATIONS

First Authorships

Farrell, S., Radford, A.D., Al Moubayed, N**.** & Noble, P.-J.M. ***PetEVAL: A veterinary free text electronic health records benchmark.*** *BioNLP @ ACL 2025 doi: pending*

Farrell, S., Anderson, K., Noble, P.-J.M. and Al Moubayed, N. (2024). ***Premature mortality analysis of 52,000 deceased cats and dogs exposes socioeconomic disparities.*** *Scientific Reports*, 14(1). doi: 10.1038/s41598-024-77385-8.

Farrell, S.*, Burton, J.*, Noble, P.-J.M. and Al Moubayed, N. (2024). ***Explainable text-tabular models for predicting mortality risk in companion animals.*** *Scientific reports, 14(1)*. doi: 10.1038/s41598-024-64551-1.

Farrell, S., Appleton, C., Noble, P.-J.M. and Al Moubayed, N. (2023). ***PetBERT: automated ICD-11 syndromic disease coding for outbreak detection in first opinion veterinary electronic health records.*** *Scientific Reports, 13(1)*. doi: 10.1038/s41598-023-45155-7.

Farrell, S., Bagcigil, A.F., Chaintoutis, S.C., Firth, C., Aydin, F.G., Hare, C., Maaland, M., Mateus, A., Vale, A.P., Windahl, U., Damborg, P., Timofte, D., Singleton, D.A. and Allerton, F. (2023). ***A multinational survey of companion animal veterinary clinicians: How can antimicrobial stewardship guidelines be optimised for the target stakeholder?*** *The Veterinary Journal*, 303, pp.106045–106045. doi: 10.1016/j.tvjl.2023.106045.

Farrell, S., McGarry, J., Noble, P.-J.M., Pinchbeck, G.J., Cantwell, S., Radford, A.D. and Singleton, D.A. (2023). ***Seasonality and other risk factors for fleas infestations in domestic dogs and cats.*** *Medical and Veterinary Entomology*, 37(2). doi: 10.1111/mve.12636.

Farrell, S., Noble, P.-J.M., Pinchbeck, G.L., Brant, B., Caravaggi, A., Singleton, D.A. and Radford, A.D. (2020). ***Seasonality and risk factors for myxomatosis in pet rabbits in Great Britain***. *Preventive Veterinary Medicine*, 176, p.104924. doi: 10.1016/j.prevetmed.2020.104924.

Co-authorships

Davies, H., Nenadic, G., Alfattni, G., Arguello Casteleiro, M., Al Moubayed, N., **Farrell, S.**, Radford, A.D. and Noble, P.-J.M. (2024). ***Text mining for disease surveillance in veterinary clinical data: part two, training computers to identify features in clinical text.*** *Frontiers in Veterinary Science*, 11. doi: 10.3389/fvets.2024.1352726.

Davies, H., Nenadic, G., Alfattni, G., Arguello Casteleiro, M., Al Moubayed, N., **Farrell, S.**, Radford, A.D. and Noble, P.-J.M. (2024). ***Text mining for disease surveillance in veterinary clinical data: part one, the language of veterinary clinical records and searching for words.*** *Frontiers in veterinary science, 11. doi: 10.3389/fvets.2024.1352239.*

Fins, I.S., Davies, H., **Farrell, S.,** Torres, J.R., Pinchbeck, G., Radford, A.D. and Noble, P.-J.M. (2023). ***Evaluating ChatGPT text mining of clinical records for companion animal obesity monitoring.*** *The Veterinary Record*, 194(3). doi: 10.1002/vetr.3669.

Smith, S.L., Anderson, E.R., Cansado-Utrilla, C., Prince, T., **Farrell, S.**, Brant, B., Smyth, S., Noble, Noble, P.-J.M., Pinchbeck, G.L., Marshall, N., Roberts, L., Hughes, G.L., Radford, A.D. and Patterson, E.I. (2021). ***SARS-CoV-2 neutralising antibodies in dogs and cats in the United Kingdom***. *Current Research in Virological Science, 2, p.100011.* doi: 10.1016/j.crviro.2021.100011.

In Review

Farrell, S., Singleton, D.A. Radford, A.D., Pinchbeck G., Noble, P.-J.M. & Al Moubayed, N**.** ***Automated Disease Classification of Veterinary Clinical Narratives for Antimicrobial Stewardship Guideline Monitoring.*** [In review]

Noble, P.-J.M., **Farrell, S.,** Al Moubayed, N**.** & Radford, A.D., ***Comprehensive representation of health-related phenotypes in one million dogs using topic modelling of electronic health records.*** [In review]

Chenghao, X., Hudson, T., Jones, B., Watson, M., **Farrell, S.**, Harmsworth-King, J., & Al Moubayed, N., **Generalizable multilingual medical text de-identification using generative instruction tuning.** [In review]

Lawson, A.,**Farrell, S**, Noble, P-J.,; Mair, T., Smith, J., Pinchbeck, G., **Quantifying and contextualising antimicrobial usage using automated disease classification tools.** [In review]

*Equal Contribution

CONFERENCES AND PRESENTATIONS

Symposium on Artificial Intelligence in Veterinary Medicine (2025) - Oral presentation: **" *Democratising Veterinary EHRs: Balancing Privacy & Open Science for the future of LLM Research in Veterinary Science*"** (Cornell University, USA) - *Best Talk Award*

Association for Veterinary Informatics Talbot Veterinary Informatics Symposium (2024) - Oral presentation: ***"Syndromic Disease Surveillance and Multi-label Classifiers for Antimicrobial Usage Assessment"*** (Virginia-Tech University, USA)

Symposium on Artificial Intelligence in Veterinary Medicine (2024) - Oral presentation: ***"PetBERT: Applications in Veterinary Syndromic Disease Surveillance"*** (Cornell University, USA)

HealTAC Annual Conference (2024) - Poster presentation: ***"Where are all the antimicrobials being used? LLM's for monitoring adherence to antimicrobial stewardship guidelines in veterinary practices"*** (Lancaster University, UK)

Annual BBSRC NLD DTP Conference (2023) - Oral presentation: ***"Disease Outbreak Detection Using Large Language Models"*** (Durham University, UK) - *Best Talk Award*

European Network for the Optimisation of Veterinary Antimicrobial Therapy (ENOVAT) Meeting (2023) - Oral presentation: ***"Survey Results on what do clinicians want from their antimicrobial stewardship guidelines"*** (University of Copenhagen, Netherlands)

HealTAC Annual Conference (2023) - Poster presentation: ***"Syndromic Disease Classification of Veterinary EHR Notes for Disease Outbreak Detection"*** (University of Manchester, UK)

Annual BBSRC NLD DTP Conference (2022) - Oral presentation: ***"Predictive Power of Large Language Models in Determining Mortality Risks in Companion Animals"*** (Durham University, UK) - *Best Talk Award*

Medical Research Foundation National PhD Conference (2022) - Poster presentation: ***"Syndromic Surveillance for Understanding Antimicrobial Usage in the veterinary community"*** (University of Bristol, UK)

European Network for the Optimisation of Veterinary Antimicrobial Therapy (ENOVAT) Meeting (2022) - Oral presentation: ***"Current Status of multinational survey of companion animal veterinary clinicians: How can antimicrobial stewardship guidelines be optimised for the target stakeholder?"*** (Aristotle University of Thessaloniki, Greece)

NGSchool Machine Learning in Computational Biology (2022) - Workshop: ***"Introduction to Natural Language Processing"*** (Warsaw, Poland)

RESEARCH SKILLS

Natural Language Processing: Successfully pretrained domain-specific models such as PetBERT, training of classifiers with minimal annotations, Retrieval Augmented Generation (RAG), topic modelling, explainability, multi-modality, named entity recognition (NER), and generative models.

Data Analysis and Statistical Modelling: Proficient in Python, R, and SQL, with extensive experience in frameworks such as Huggingface Transformers, scikit-learn, and PyTorch. Skilled in designing scalable pipelines optimized for low-resource environments and managing the analysis of very large datasets (10M+ records). Proficient in Git-based version control and collaborative project management. Experienced in hypothesis testing, regression analysis, and advanced statistical modelling

Data Visualisation: Adept at creating clear, informative, and publication-quality visualisations using Matplotlib, Seaborn, ggplot2, and Plotly. Skilled in website creation for serving models and presenting visualisations in an interactive manner

Teaching and Mentoring: Experienced in teaching NLP to master's-level students, adapting methods for varied skill levels. Mentoring a master's students on an epidemiology-based project. Pursuing an Associate Fellowship in Higher Education to refine my teaching practice.

Communication and Presentations: Experienced in delivering technical presentations to limited prior knowledge audiences. Capable of tailoring complex concepts to both expert and lay audiences, ensuring clarity and engagement. Won best talk awards.

Ethical and Scientific Writing: Published in peer-reviewed journals, skilled at structuring clear and concise papers, emphasising reproducibility and scientific rigor. Experience in writing ethics proposals.

EMPLOYMENT

Data Science Intern

Evergreen Life

- Developed a Retrieval Augmented Generation pipeline utilising generative language model (LLM) to deliver personalised healthcare advice, ensuring information safety and accuracy by aligning outputs with the Evergreen Life article repository.
- Designed an algorithm-driven content recommendation system, tailoring advice to individual user needs.
- My work is being integrated into an app serving over 1 million NHS patients, enhancing its capability to provide personalised, reliable healthcare guidance.

Natural Language Processing Demonstrator
January 2021 – Present
Durham University

- Delivered comprehensive NLP curriculum from foundational machine learning to advanced Deep Learning and Transformers, adapting technical content for diverse audiences ranging from MSc Computer Science to MBA Business Analytics students

Undergraduate Researcher
July 2020 – September 2021
University of Liverpool
- Conducted research on geospatial and risk factors for fleas as a second-year undergraduate
- Utilised electronic health records (EHRs) from over 34,000 animals across UK first-opinion veterinary practices.
- Published findings in the Journal of Medical and Veterinary Entomology

Undergraduate Researcher
June 2019 – January 2020
University of Liverpool
- Conducted research on geospatial and risk factors for Myxomatosis as a first-year undergraduate
- First exposure to large-scale data analytics, applying statistical methodologies such as multivariate logistic modelling for risk factor analysis
- Secured funding to produce and distribute educational posters summarising research findings to veterinary practices across the UK.
- Published findings in the Journal of Preventive Veterinary Medicine

Customer Experience Supervisor
October 2016 – September 2021
Sainsbury's

FINDINGS AND AWARDS

- **Creating a webpage hosting the Archive of Protocols for Veterinary Microbiology Investigations** – EU COST Action Virtual Mobility grant (8/4/2024) - €1500
- **Drafting and Launching of the Equine Antimicrobial Treatment Survey** – EU COST Action Virtual Mobility grant (01/09/2022) - €1500
- **Refinement, redrafting and relaunching the veterinary antimicrobial treatment survey**– EU COST Action Virtual Mobility grant (01/11/2021) - €1500

TEACHING

Natural Language Processing Demonstrator
January 2021 – Present
- Leading workshops from statistical methods for natural language processing to large language model training
- Designed workshop materials including guides and code
- Taught to the taught-Masters Computer Science Programme

Natural Language Analysis Demonstrator
January 2021 – Present
- Leading workshops from statistical methods for natural language processing to large language model training
- Taught to the Master of Business Administration (MBA) programme

- Limited prior knowledge in python, therefore classes are much more focussed on understanding the code and attaining results before grasping with technical knowledge around deep learning methodologies

REFERENCES

Dr Noura Al Moubayed
Associate Professor in Computer Science
Department of Computer Science, Durham University, UK
noura.al-moubayed@durham.ac.uk

Professor Peter-John Noble
Professor in Small Animal Internal Medicine
Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, UK
Rtnorle@liverpool.ac.uk

Dr James Harmsworth-King
Chief Medical Officer
Evergreen Life, Manchester, UK
James.Harmsworth-King@evergreen-life.co.uk

Professor Alan Radford
Professor in Veterinary Health Informatics
Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, UK
A.D.Radford@liverpool.ac.uk