Wenrui Fan

wenrui.fan@sheffield.ac.uk +44 (0) 7572 857388 wenruifan.github.io wenruifan
Google Scholar Web of Science LinkedIn Hugging Face ORCID: 0009-0007-1394-0092

Education

- **PhD Candidate in Computer Science**
  - University of Sheffield, UK 01.2024 – Present
- **MSc in Robotics**
  - University of Sheffield, UK; Distinction 09.2021 – 09.2022
- **BSc in Flight Vehicle Design and Engineering**
  - Beijing Institute of Technology, China; Distinction BSc thesis 09.2017 – 07.2021

Appointments

- **AI Research Engineer – Centre for Machine Intelligence, University of Sheffield** 09.2023 – Present
  - Lead and contribute to multimodal AI research in neuroimaging, precision medicine, cancer, protein science, and molecular modelling.
  - Develop models, datasets, evaluation protocols, interpretation methods, and reproducible research software.
  - Coordinate technical work with clinical, biomedical, and interdisciplinary collaborators.
  - Contribute to collaborative funding proposals and open research infrastructure.
- **Research Engineer, AI for Manufacturing – Chongqing Research Institute of HIT** 11.2022 – 07.2023
  - Led machine-learning development for semiconductor wafer anomaly detection.
  - Improved detection performance and computational efficiency.
  - Contributed to foundation-model architectures for intelligent manufacturing.

Technical Skills

- **AI and machine learning:** Multimodal learning, scientific AI, foundation-model adaptation, vision–language pre-training, graph learning, agentic AI, and generative modelling.
- **Evaluation and analysis:** Baseline design, independent-cohort validation, subgroup analysis, model interpretation, causal analysis, and benchmarking.
- **Research software engineering:** Python, PyTorch ecosystem, scikit-learn, Hugging Face, Git and GitHub, Linux, Docker, automated testing, continuous integration, Python packaging, and FAIR workflows.

Selected Research Projects

- **Neuropathic Pain Drug-response Prediction** – Code
  - Led the design and PyTorch implementation of a multimodal framework using resting-state fMRI time-series and functional-connectivity modalities.
  - Adapted a foundation model pre-trained on UK Biobank data to small clinical cohorts.
  - Designed baselines and evaluation across two clinical datasets.
  - Conducted subgroup-specific interpretation and compared learned patterns with medical and biological findings and theories.
- **MeDSLIP: Medical Vision–Language Pre-training** – Code | Model
  - Led model design, implementation, experimentation, and release for fine-grained medical image–text alignment.
  - Evaluated zero-shot learning, transfer learning, and finetuning.
  - Released the code and pre-trained model for external reproduction and reuse.
- **Parkinson's Disease Risk-factor Analysis**
  - Led the machine-learning work for a UK Biobank study of potential Parkinson's disease risk factors.
  - Combined predictive modelling, age- and sex-stratified analysis, model interpretation, and causal inference.
  - Distinguished predictive contribution from causal evidence to support appropriately cautious conclusions.
- **AI-driven Multistate Protein Design**
  - Develop data infrastructure integrating protein sequence, structure, state, transition, and functional information.
  - Construct benchmark tasks, datasets, splits, baselines, and evaluation metrics for multistate structure prediction and inverse folding.
  - Develop state-aware annotation, representation-learning, and generative modelling methods for multistate protein design.
- **Selective Anticancer Peptide Design**
  - Contribute to de novo peptide generation, selective anticancer screening, and cancer-type prediction.
  - Design additional computational experiments and support evaluation of generated candidates.

Open-source Software and Data Infrastructure

- **KaleProtein (v0.1.0a1, PyPI)** – Founder and lead developer
  - Own the software architecture, implementation, testing, documentation, packaging, versioning, and release workflow.
  - Support protein scientists through standardised model and dataset registration and loading interfaces.
  - Provide runnable DrugBAN and MapDiff workflows with the required datasets.
  - Maintain automated testing and continuous integration through GitHub Actions.
- **PyKale (480+ GitHub stars; 34K+ downloads)** – Core contributor and maintainer
  - Contribute to a multimodal machine-learning library included in the PyTorch Ecosystem.
  - Develop meta-learning modules, tests, documentation, and an end-to-end cancer multi-omics tutorial.
  - Support repository maintenance and a user-driven redesign informed by surveys and semi-structured interviews.
  - Fourth-largest contributor by commit count after its three co-founders.
- **MuSProt** – Lead developer
  - Constructed an open, million-scale multistate protein database.
  - Developed and deployed a Hugging Face-hosted web platform, providing search, visualisation, and data download.
  - Released the reproducible dataset-construction workflow, including data collection and automated annotation methods.

Publications

*Equal contribution.

Conference Papers

[C1] **W. Fan**, L.M.R. Rizky, J. Zhang, C. Chen, H. Lu, K. Teh, D. Selvarajah, and S. Zhou, "Foundation-Model-boosted multimodal learning for fMRI-based neuropathic pain drug response prediction", in *the 28th Int. Conf. on Medical Image Computing and Computer Assisted Intervention (**MICCAI**,* **CORE A***)*, pp. 238–248, Sep. 2025. DOI: 10.1007/978-3-032-05182-0_24.

[C2] J. Zhang, X. Liu, W. Wu, S. Tabakhi, **W. Fan**, S. Zhou, K.L. Tee, T.S. Wong, and H. Lu, "Classifying the Stoichiometry of Virus-like Particles with Interpretable Machine Learning", in *47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (**EMBC**)*, pp. 1–7, Jul. 2025. DOI: 10.1109/EMBC58623.2025.11254123.

[C3] M.N.I. Suvon, P.C. Tripathi, **W. Fan**, S. Zhou, X. Liu, S. Alabed, V. Osmani, A. Swift, C. Chen, and H. Lu, "Multimodal Variational Autoencoder for Low-cost Cardiac Hemodynamics Instability Detection", in *the 27th Int. Conf. on Medical Image Computing and Computer Assisted Intervention (**MICCAI**,* **CORE A***)*, pp. 296–306, Oct. 2024. DOI: 10.1007/978-3-031-72378-0_28.

[C4] J. Wang, B. Chen, **W. Fan**, and Y. Liu, "Co-salient object detection via discriminative prototypes contrast", in *the 2024 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (**ICASSP**)*, pp. 7390–7394, Apr. 2024. DOI: 10.1109/ICASSP48485.2024.10446685.

[C5] **W. Fan**, "Path planning for mobile robot in 3D environment based on ant colony algorithm", in *the 2nd Int. Conf. on Artificial Intelligence and Information Systems (**ICAIIS**)*, p. 012095, May. 2021. DOI: 10.1088/1742-6596/1982/1/012095.

Journal Papers

[J1] P. Bai, X. Liu, **W. Fan**, T. Jiang, S. Alabed, W.K. Cheung, and H. Lu, "Geometry-aware Line Graph Transformer Pre-training for Molecular Property Prediction", ***IEEE Transactions on Neural Networks and Learning Systems (IF: 8.9)***, May. 2026.

[J2] T. Payne*, **W. Fan***, C. Chen, S. Zhou, V. Osmani, H. Lu, and O. Bandmann, "Risk of Parkinson's disease following cholecystectomy: a prospective cohort study in the UK Biobank," ***npj Parkinson's Disease (IF: 8.2)***. Under revision.

[J3] M. N. I. Suvon, S. Zhou, P. C. Tripathi, **W. Fan**, S. Alabed, B. Khanal, V. Osmani, A. J. Swift, C. Chen and H. Lu, "Multimodal Latent Fusion of ECG Leads for Early Assessment of Pulmonary Hypertension," ***IEEE Journal of Biomedical and Health Informatics (IF: 6.7)***. arXiv: 2503.13470, Under review after major revision.

Preprints & Abstracts

[P1] **W. Fan**, M. N. I. Suvon, S. Zhou, X. Liu, S. Alabed, V. Osmani, A. Swift, C. Chen, & H. Lu, "MeDSLIP: Medical Dual-Stream Language-Image Pre-training with Pathology-Anatomy Semantic Alignment", *arXiv: 2403.10635*, 2024.

Leadership

- **Scientific AI Benchmarking Spinout** – Entrepreneur Lead
  - Lead the development of a prospective University of Sheffield spinout focused on deployment-centric benchmarking and qualification of scientific AI systems.
  - Define the technical and commercial proposition, translating real-world deployment conditions, constraints, failure modes, robustness, uncertainty, and generalisation risks into measurable qualification tests, reusable tools, and standardised workflows.
  - Coordinate the spinout team with Prof. Haiping Lu and Dr Xianyuan Liu, leading commercialisation planning, customer discovery, IP preparation and take commercialization training of University of Sheffield.
- **Technical and research project leadership**
  - Translate biomedical questions into tractable machine-learning tasks and make key methodological decisions in the neuropathic pain and Parkinson's disease projects.
  - Coordinate technical delivery, collaborator discussions, action tracking, and follow-up.
  - Independently manage the end-to-end development and release of KaleProtein.

- **Co-organized events**
  - PyKale tutorial at IEEE EMBC 2025, Copenhagen, Denmark, July 2025.
  - Fourth Workshop on Multimodal AI, London, UK, June 2026.
  - Second Mini-Hackathon on Multimodal AI, London, UK, June 2026.
  - Second Multimodal AI Community Forum, Sheffield, UK, March 2026.
  - Third Workshop on Multimodal AI, London, UK, September 2025.
  - First Mini-Hackathon on Multimodal AI, London, UK, September 2025.
  - Second Workshop on Multimodal AI, Sheffield, London, UK, September 2024.
- **Bio-lab research assistant hackathon team – Team lead; Second place**
  - Built the main agentic system for laboratory management, experiment planning, parameter calculation, and real-time voice question answering.
  - Integrated supporting tools developed by other members of the team.
- **Self-driving protein and materials design lab hackathon team – Team lead; Second place**
  - Led a separate interdisciplinary team of biology, chemistry, mathematics, active-learning, and AI researchers.
  - Designed and implemented an agentic workflow for requirement-driven metal-binding protein design, evaluation, screening, and optimisation.
  - Led the integration, demonstration design, and technical presentation of the prototype.

Selected Grant Applications

- **UK Sovereign AI Strategic Assets proposal – Co-Investigator; approximately £1.27M**
  - Originated the core concept for an AI-ready benchmarking dataset spanning proteins and organic materials.
  - Led most of the expression-of-interest development and coordinated technical follow-up with academic and industry partners.
- **ARIA Activation Partner proposal – Co-applicant; £6.39M**
  - Co-developed the project concept, technical approach, and delivery plan.
- **UKRI Generative AI Hub proposals – Co-applicant; two proposals of £100K each**
  - Led the concept, technical planning, and application draft for one proposal.
  - Contributed to the conception of a second proposal.
- **Innovate UK Frontier AI Benchmarking Datasets proposals – Co-applicant; two proposals of £750K each**
  - Contributed technical concepts and benchmarking methodology.
- **ARIA x Encode AI Fellowship – Applicant; £115K**
  - Independently developed an application on AI-driven multistate protein design.

Dissemination, Training, and Professional Service

- **Selected Presentations**
  - Neuropathic pain drug-response prediction, MICCAI, Daejeon, September 2025.
  - Poster and one-minute pitch: “State-Aware Functional Annotations for Multi-State Design of Dynamic Proteins”, AI, Engineering Biology and Beyond, University of Bristol, January 2026.
  - Poster on multistate protein research, Prague Protein Spring, Prague, 2026.
  - “AI for Parkinson’s Disease”, Sheffield AI Community Meeting, University of Sheffield, March 2024.
- **Selected Journal Reviewer**
  - IEEE Transactions on Neural Networks and Learning Systems.
  - IEEE Journal of Biomedical and Health Informatics.
  - Computational and Structural Biotechnology Journal.
- **Selected Conference Reviewer**
  - International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2026.
  - MICCAI Workshop on Advancing Data Solutions in Medical Imaging AI (ADSMI), 2024.