

Outline of Research & Development Projects: Soumya Banerjee

1. Computational Immunology & Within-Host Pathogen Dynamics

A. Modelling within-host dynamics of viral infections (West Nile virus)

- Timeframe: PhD & postdoc (2009–2016)
- Role: Lead model developer / first author
- Goals: Quantitatively describe within-host replication and immune response dynamics across host body sizes and time scales.
- Methods / Tools: ODEs, hierarchical Bayesian inference, stochastic modelling, parameter estimation, MATLAB/Python.
- Key outputs: Multiple J. R. Soc. Interface papers (e.g., 2016, 2017); validated against experimental murine infection data.
- Impact / Next steps: Provides mechanistic insights for pathogen control; potential extension into Bayesian hierarchical models for multi-cohort clinical data.

B. Immune-system inspired distributed search & response (RADAR, Modular RADAR)

- Timeframe: PhD & early postdoc (2009–2011)
 - Role: Co-author / algorithm designer
 - Goals: Design scalable, decentralized search/response algorithms inspired by immune system processes for distributed computing and resource allocation.
 - Methods / Tools: Agent-based modelling, multi-agent systems, simulation in Python/C++.
 - Key outputs: ICARIS / ALife publications; conceptual architecture for robust decentralized systems.
 - Impact / Next steps: Apply to modern distributed systems (edge computing, sensor networks) and evaluate with real-world workloads.
-

2. Systems & Cell Biology — Signalling and Dynamics

A. Rho GTPase signalling and subcellular contraction dynamics

- Timeframe: Max Planck (2013–2014); Cell Reports / J. Cell Biology collaborations (2017–2020)
- Role: Modeler and computational collaborator
- Goals: Explain spatiotemporal patterns of cell contraction and the role of Rho amplification.
- Methods / Tools: Non-linear ODEs, dynamical systems, hierarchical Bayesian models, automated image analysis (CellProfiler, ImageJ), MATLAB/Python.
- Key outputs: J. Cell Biology, Cell Reports papers; optogenetic model validation.
- Impact / Next steps: Integrate with single-cell datasets and predictive control strategies for synthetic biology applications.

B. Bioorthogonal small-molecule switch systems (collaboration)

- Timeframe: 2014
 - Role: Computational support and model validation
 - Goals / outputs: Contributed to experimental design and quantitative interpretation; publication in Angewandte Chemie.
-

3. Microbiome, Metabolomics & Clinical Translational Projects

A. Immune–microbiome interactions in Inflammatory Bowel Disease (IBD)

- Timeframe: Harvard/Broad (2014–2015) and subsequent collaborations
- Role: Computational analyst, modeler
- Goals: Identify microbial metabolites and networks influencing host immune responses in IBD.
- Methods / Tools: Metagenomic/metabolomic predictive models, statistical association, Python/R pipelines.

- Key outputs: Contributions to Nature Communications and Gut; identified candidate microbial compounds and IL-1/IL-23 networks.
 - Impact / Next steps: Translate findings into biomarker panels or therapeutic targets; integrate multi-omics for patient stratification.
-

4. Explainable & Interpretable AI for Healthcare

A. Class-contrastive, human-interpretable ML for mortality prediction in severe mental illness

- Timeframe: Cambridge postdoc / Senior Research Fellow (2019–2022; 2021 Nature Partner Journal paper)
- Role: Lead developer / first author
- Goals: Build interpretable models that clinicians can understand and trust to predict mortality risk.
- Methods / Tools: Class-contrastive explanations, feature attribution, structured EHR preprocessing, Python, reproducible pipelines.
- Key outputs: NPJ Schizophrenia (2021); follow-up explanations & software (2024 Life paper); patient involvement studies (Patterns, 2022).
- Impact / Next steps: Deploy prototypes in clinical workflows (pilot evaluations); combine with federated analysis approaches for privacy.

B. Generating complex explanations for AI models applied to clinical data

- Timeframe: 2022–2024
- Role: Principal investigator / author
- Goals: Develop methods to produce clinically actionable, human-readable explanations from complex ML models.
- Methods / Tools: Model-agnostic explanation methods, R/Python, visualisations, human-centred design and PPI (patient and public involvement).
- Key outputs: Life (2024); Patterns (2022) framework for patient/public involvement.

5. Privacy-Preserving & Federated Analysis (DataSHIELD Ecosystem)

A. Development of privacy-preserving survival analysis tools (dsSurvival, dsSurvival2.0)

- Timeframe: 2022–2025
- Role: Developer / maintainer / first author on related notes
- Goals: Enable non-disclosive survival modelling across distributed clinical cohorts.
- Methods / Tools: R, DataSHIELD platform, Shiny apps (ShinyDataSHIELD), statistical survival methods adapted for federated settings.
- Key outputs: dsSurvival packages, BMC Research Notes (2022; 2023), International Journal of Epidemiology ShinyDataSHIELD application (2022), DataSHIELD collaboration (2025 Bioinformatics Advances paper).
- Impact / Next steps: Broaden adoption across consortia, integrate with synthetic data (dsSynthetic) and differential privacy techniques.

B. Synthetic data & non-disclosive tooling (dsSynthetic, ShinyDataSHIELD)

- Timeframe: 2022–2023
- Role: Author / software lead
- Goals: Provide synthetic data generators and GUI tools to lower barriers to federated analysis.
- Key outputs: dsSynthetic (BMC Research Notes 2022), ShinyDataSHIELD application.

6. Socio-economic & Complex Systems Modelling

A. Crime dynamics & law-enforcement modelling

- Timeframe: CSIRO / earlier postdoc (2015–2016)
- Role: Modeler / co-author

- Goals: Explain macro-scaling laws of crime through competitive dynamics models between criminals and law enforcement.
- Methods / Tools: Agent-based models, statistical analysis of open data, R/Python.
- Key outputs: Palgrave Communications (2015); policy-relevant insights into resource allocation for policing.

B. Transport and infrastructure demand modelling

- Timeframe: 2016
 - Role: Co-investigator
 - Goals / outputs: Multiple conference papers on intermodal terminal demand and forecasting in big-data contexts.
-

7. Machine Intelligence, Abstraction & AI Theory

A. Neural networks for abstraction & reasoning; ARC work

- Timeframe: 2024–2025
- Role: Co-author / contributor
- Goals: Investigate architectures for abstraction and reasoning (Abstraction and Reasoning Corpus), hybrid human–AI interaction strategies.
- Key outputs: Scientific Reports (2024); workshop/conference contributions (2025).
- Impact / Next steps: Bridge symbolic and subsymbolic methods; build interactive tools that augment human reasoning.

B. Conceptual work on intelligence across humans, animals & machines

- Timeframe: 2024–2025
- Role: Author / conceptual lead
- Goals: Reframe intelligence debates with interdisciplinary perspectives (philosophy, comparative cognition, AI ethics).

- Key outputs: AI & Society (2024), 2025 workshop papers.
-

8. Software, Reproducibility & Teaching-Led R&D

A. Reproducible research tools and teaching materials

- Timeframe: Ongoing (2017–2025)
- Role: Author / instructor
- Goals: Produce open, well-documented teaching materials and reproducible pipelines for ML and data visualisation.
- Methods / Tools: R, Python, GitHub repos (teaching_reproducible_science_R, visualization_lecture), Git/CI practices.
- Key outputs: Public course materials, workshops, adoption by students/colleagues.

All key teaching material I have developed are available here:

<https://sites.google.com/site/neelsoumya/teaching>

An example of a course that I have developed and taught is here:

<https://cambiotraining.github.io/ml-unsupervised/>

B. Open-source packages and R tools

- Timeframe: 2022–2025
- Role: Developer / maintainer
- Goals: Create practical packages for federated survival analysis and synthetic data (dsSurvival, dsSurvivalClient, dsSynthetic).
- Key outputs: GitHub repositories, publications describing the software, Shiny apps.

9. Cross-cutting & Translational Activities

- Patient and public involvement (PPI): Frameworks and case studies for involving patients in AI research to build trust (Patterns, 2022). Central to multiple projects in explainable AI and clinical translation.
 - Funding & infrastructure: Involved in AI@CAM pilot grants and OpenAI Researcher Access; experience writing grants and budgets.
 - Supervision & capacity building: Supervision of MPhil/PhD students and development of curricula linking research to student projects.
-

10. Suggested Packaging & Uses of This Outline

1. Grant applications: Use the thematic project descriptions as project briefs; highlight outputs, collaborators, and planned next steps.
 2. Job/selection materials: Convert each theme into short project bullets (2–3 lines) with 1–2 highlighted outputs (paper/package/grant).
 3. Website / public profiles: Expand each project with 1–2 visuals (figures, pipeline diagrams) and links to code and publications.
-

11. Quick inventory: Select key publications & software linked to projects

- Explainable AI / Clinical ML: NPJ Schizophrenia (2021); Life (2024); Patterns (2022)
- Federated analytics / DataSHIELD: dsSurvival, dsSynthetic, ShinyDataSHIELD; BMC Research Notes and Int J Epidemiol papers (2022–2025)
- Systems & immunology: Multiple J. R. Soc. Interface, J. Cell Biology, Nature Communications, Gut papers

- Software: Reproducible research repos; dsSurvival family on GitHub

<https://github.com/neelsoumya/dsSurvival>
