Transforming Cancer Care Through Artificial Intelligence and Machine Learning: Insights from USA and Australia

Dr Maryam Arshad
Churchill Fellow 2024

the CHURCHILL fellowship

KING'S College LONDON

NHS
Guy's and St Thomas'
NHS Foundation Trust

THE FRANCIS CRICK INSTITUTE

Contents

Acknowledgement

I am deeply grateful to the Winston Churchill Memorial Trust for awarding me this Fellowship and for enabling the international learning that shaped this report. Their support provided a unique opportunity to explore how artificial intelligence can be responsibly integrated into cancer research and healthcare systems.

My sincere thanks go to **Dr Alexander (Sasha) Gusev** and the members of the Gusev Laboratory at the Dana-Farber Cancer Institute and Harvard Medical School. I am particularly grateful for their generosity in sharing their research, methods and time, and for the thoughtful discussions that significantly strengthened my understanding of quantitative genomics, multi-omic integration and machine-learning approaches in oncology.

I wish to extend my appreciation to **Professor Lyle J. Palmer** and his team at the **Australian Institute for Machine Learning (AIML)** and the University of Adelaide. Their open, collaborative approach provided invaluable insight into applied machine learning, population-health analytics and implementation science. I am especially thankful for the opportunity to observe the practical realities of interdisciplinary AI development within a mature translational ecosystem.

I am indebted to colleagues at **King's College London**, whose mentorship and support throughout my Fellowship have been instrumental. I also acknowledge the clinicians, researchers and healthcare professionals across the NHS who informed my understanding of the challenges and opportunities for AI adoption within the UK.

Finally, I thank my family, friends and professional mentors for their encouragement throughout this journey. Their support enabled me to undertake this Fellowship with purpose, curiosity and ambition.

To all who contributed in ways large and small: **thank you**. This report reflects the generosity, expertise and collaboration of many individuals across three continents.

1. Executive Summary

I am a cancer researcher and clinician–scientist specialising in the tumour microenvironment, immuno-oncology, and translational data science. As a Senior Research Fellow at King's College London, my work focuses on developing AI-enabled approaches for cancer risk stratification, early detection, and treatment decision support. I applied for the Churchill Fellowship to understand how world-leading centres operationalise artificial intelligence (AI) and machine learning (ML) in real healthcare settings, and to identify strategies that could meaningfully accelerate responsible AI adoption across the UK, particularly within the NHS and cancer research ecosystem.

In line with these aims, my Fellowship began in the USA (**Boston)**, where I spent time at the **Dana-Farber Cancer Institute (DFCI)** and the **Gusev Lab**, one of the foremost groups in computational cancer genomics. This visit offered direct insight into how advanced ML tools such as OncoNPC for cancers of unknown primary (CUP), polygenic risk score modelling, and multi-omic integration frameworks are being developed, validated, and translated into precision oncology (1). Observing the Gusev Lab's interdisciplinary culture and rigorous methodological standards clarified how AI can support clinical questions in oncology, from improving diagnostic certainty to enabling more personalised treatment decisions (2).

I then travelled to **Australia**, where I undertook a structured placement at the **Australian Institute for Machine Learning (AIML)** and shadowed **Professor Lyle J. Palmer**, an international leader in genetic epidemiology, public-health analytics and medical ML. AIML provided a contrasting but complementary perspective to Boston: a mature translational ecosystem where AI solutions are codesigned with clinicians, embedded into real workflows, and evaluated through national-level data infrastructures and implementation-science methods. At AIML, I observed ML tools being applied to imaging, chronic disease prognosis, biobank-scale modelling, and public-health platforms each developed through a clinically driven, data-ready, interdisciplinary approach (3,4).

Across both international visits, a consistent set of insights emerged. High impact AI in healthcare does not begin with algorithms; it begins with well-defined clinical problems, high quality linked datasets, interdisciplinary teams, and governance frameworks that ensure fairness, safety, and public trust. Equally important is rigorous external validation and thoughtful implementation planning both of which determine whether AI systems succeed in real clinical environments. These lessons reaffirmed that the UK's opportunity lies not only in scientific innovation but in building the infrastructure, standards, and workforce necessary to translate AI into routine practice.

Key findings from the Fellowship include:

- **AI delivers the greatest value when developed in direct response to clearly articulated clinical needs**, such as cancer of unknown primary (CUP) classification, imaging backlogs, chronic-disease risk prediction or personalised oncology.
- **Data quality, interoperability and linkage are foundational**, enabling ML models to learn from diverse, representative, and clinically meaningful datasets.
- **Robust evaluation including external and temporal validation, subgroup analysis and real-world testing is essential** for safe and equitable deployment.
- **Interdisciplinary collaboration between clinicians, data scientists, epidemiologists, geneticists and engineers are a prerequisite for translational progress.**
- **Ethics, fairness, transparency, and patient trust must underpin every stage of AI development and implementation**, especially in high-stakes fields such as oncology.

These insights directly strengthen my ongoing research at King's and my leadership in AI-enabled cancer prediction and decision support. This Fellowship has also enabled me to secure new international collaborations, apply for competitive AI funding, and present my early outputs at the **ESMO AI in Oncology Conference**, positioning the UK as an active contributor in the global AI-oncology landscape.

This report presents the detailed findings of my visits to Boston and Australia and offers a set of strategic recommendations for UK policymakers, NHS organisations, academic institutions, clinicians, and AI developers. It outlines how the UK can build a trustworthy, clinically driven infrastructure for AI adoption ensuring that emerging technologies genuinely improve outcomes for patients and support the future of precision medicine.

2. Aims of the Fellowship

The overarching aim of my Churchill Fellowship was to explore how artificial intelligence (AI) and machine learning (ML) can be harnessed to improve healthcare delivery, with particular emphasis on cancer research, precision medicine, and clinical decision-making. Central to this objective was understanding how internationally leading centres design, validate and integrate AI-driven tools into real-world clinical workflows, and identifying lessons that could be applied within the UK especially the NHS.

To achieve this, the Fellowship focused on the following aims:

2.1 Investigate global best practice in AI and ML for healthcare

This aim centred on examining how world-leading institutions particularly those in Australia develop and deploy advanced ML systems across fields such as clinical imaging, genomics, population health and risk prediction. It involved understanding the scientific, technical, and organisational frameworks that support successful AI adoption.

2.2 Explore the application of AI/ML to cancer research and precision medicine

This component focused on assessing how AI is being used to identify cancer-associated genes, decode tumour biology, model therapeutic responses, and guide personalised treatment decisions. Given my research in tumour immunology and the tumour microenvironment, I sought to understand how ML can accelerate discovery and improve patient outcomes.

2.3 Learn from interdisciplinary teams implementing AI in clinical practice

A core aim was to observe how clinicians, computer scientists, epidemiologists, engineers, and public-health experts collaborate to create AI tools that are clinically meaningful, reliable, and safe. This included exploring how interdisciplinary structures operate in practice, and how such models might be adapted for UK clinical and research environments.

2.4 Examine data infrastructure, governance, and ethics frameworks

This objective focused on understanding how leading centres manage data quality, interoperability, patient privacy and model governance elements essential for responsible and equitable use of AI. I also examined how institutions address issues such as algorithmic bias, fairness, transparency, and public trust.

2.5 Analyse real-world implementation pathways for AI tools

A key part of the Fellowship was studying how institutions move from algorithm development to clinical integration. This included analysing validation processes, workflow redesign, training requirements, evaluation strategies, stakeholder engagement and long-term monitoring elements critical for safe and sustainable deployment.

2.6 Identify actionable lessons for the NHS and UK research ecosystem

The Fellowship aimed to gather insights that could inform future UK strategies for AI-enabled healthcare, including priorities such as national data infrastructure, regulatory

requirements, workforce skills, research–clinical partnerships and technological adoption frameworks.

2.7 Strengthen international partnerships in AI-based healthcare innovation

Finally, the Fellowship sought to build lasting relationships with international leaders including Professor Lyle J. Palmer and the Australian Institute for Machine Learning (AIML) to support ongoing collaboration between global AI research groups and UK academic and clinical institutions.

3. Background and Rationale

Artificial intelligence (AI) and machine learning (ML) are transforming modern healthcare, particularly in oncology, genomics, and clinical diagnostics. These technologies offer powerful opportunities for earlier detection, personalised treatment planning and more efficient decision-making (5). Yet their translation into real-world clinical settings remains uneven. Many promising models never progress beyond research prototypes, while health systems struggle with fragmented data infrastructures, unclear implementation pathways and variable governance standards. My Churchill Fellowship was established to examine how leading international centres address these challenges and to identify lessons that could strengthen AI adoption within the UK especially across the NHS.

My research at King's College London focuses on the tumour microenvironment, immune responses, and the use of advanced computational approaches to analyse large, complex biological datasets. With the growth of genomic sequencing, digital pathology and multimodal imaging, cancer research has become increasingly data-rich. However, much of the analytic potential of these datasets cannot be realised without robust ML methods and the system-level capacity to implement them safely. The NHS, with its extensive clinical records and national genomic infrastructure, is uniquely positioned to benefit from AI but only if the underlying systems, governance, and workforce are equipped to support translation (6).

Australia was selected as a key site for the Fellowship because of its strong international reputation in applied ML, particularly through the AIML. AIML's focus on clinically grounded AI, its integration with public-health agencies and its national data-linkage capabilities provides a model for how AI can be embedded into a healthcare system with challenges similar to the UK's (7). Shadowing Professor Lyle J. Palmer a leader in genetic epidemiology, chronic-disease modelling and population-scale data science offered a unique opportunity to observe how interdisciplinary teams design, evaluate and implement AI tools that are both scientifically robust and operationally feasible (8).

The Boston component of the Fellowship, undertaken at the Dana-Farber Cancer Institute and Harvard Medical School, provided a complementary perspective by offering exposure to advanced machine-learning approaches in precision oncology, multi-omic data integration and the development of state-of-the-art computational tools such as OncoNPC. This experience allowed the Fellowship to encompass two contrasting yet highly effective models of AI-driven innovation: the Australian model, centred on population-health analytics,

data-linkage infrastructure and system-level implementation, and the Boston model, characterised by its focus on computational genomics, quantitative cancer science and translational precision medicine (9). Together, these environments offered a comparative framework for understanding how different institutional ecosystems enable the design, validation, and deployment of clinically impactful AI technologies.

These international examples helped me explore three core questions underpinning the Fellowship:

1. **What scientific and infrastructural conditions enable successful AI adoption in healthcare?**
2. **How do leading centres develop, validate, and implement AI tools that are clinically meaningful and trustworthy?**
3. **What lessons can the UK apply to ensure safe, equitable and scalable integration of AI into the NHS?**

The rationale for the Fellowship was further shaped by several urgent needs within UK healthcare:

- **Robust implementation pathways:** Many UK AI models stall at the proof-of-concept stage without clear routes into clinical use.
- **Data quality and interoperability:** NHS datasets are extensive but inconsistent, limiting model performance and equity.
- **Interdisciplinary capacity:** Effective AI requires coordinated input from clinicians, data scientists, engineers, ethicists, and implementation expert's structures still developing in the UK.
- **Rising demand for precision oncology:** Increasing reliance on genomics, digital pathology and predictive tools requires advanced computational infrastructure.
- **Ethical and equitable governance:** Public trust depends on transparency, fairness assessment, and rigorous oversight across the AI lifecycle.

By examining how Australia and Boston address these issues through mature data ecosystems, interdisciplinary team science and rigorous evaluation frameworks this Fellowship aimed to identify practical strategies that can inform UK policy, research investment and clinical integration (10,11).

Ultimately, the rationale for this work rests on a simple premise: AI has the potential to improve outcomes, reduce diagnostic delays and support personalised cancer care, but only when developed and deployed within systems designed for safety, reproducibility, and equity (12). This Fellowship provided the opportunity to study those systems in practice and to generate insights that can help the UK realise the full potential of AI-enabled healthcare.

4. Quantitative Genomics and Machine Learning: Visit to the Gusev Lab, Dana-Farber Cancer Institute & Harvard Medical School

4.1. Background and Purpose of the Visit

As part of my Churchill Fellowship examining the translation of AI and ML into healthcare, I visited the laboratory of Dr Alexander ("Sasha") Gusev at the Dana-Farber Cancer Institute and Harvard Medical School. Dr Gusev, a statistical geneticist and Associate Professor of Medicine, leads a programme focused on developing computational and ML approaches to characterise genetic architecture, tumour biology and therapeutic response in cancer.

This visit was designed to complement my time in Australia by providing exposure to a leading US centre of computational and translational oncology, with a particular emphasis on quantitative genetics, multi-omic integration and ML-based predictive modelling. Together, the two components of the Fellowship offered contrasting yet complementary perspectives on AI innovation: a population-health, system-level model in Australia and a precision-oncology, genomics-centred model in Boston.

The specific aims of the Boston visit were to:

- Examine how AI/ML is applied in quantitative genetics and cancer-risk research.
- Learn about emerging predictive-modelling techniques for cancer risk assessment and therapy response.
- Understand strategies for addressing data-quality issues, algorithmic bias and the challenges of clinical integration.
- Observe interdisciplinary collaboration between geneticists, oncologists, data scientists and computational biologists.
- Explore how such approaches could improve diagnostics and patient outcomes in settings analogous to the NHS.

4.2. Overview of Dr Gusev's Research Programme

Dr Gusev's group develops methods at the interface of statistical genetics, cancer genomics and ML, with several recurring themes:

1. **Machine learning classifiers for tumour type and origin**
2. **Polygenic risk scores (PRS) and germline–somatic interactions**
3. **Integrative multi-omics to map causal mechanisms and therapeutic targets**
4. **Bias, ancestry and generalisability in genetic and ML models**

4.2.1 OncoNPC: ML for Cancers of Unknown Primary

A central highlight of my visit was the opportunity to study OncoNPC, a machine-learning classifier designed to predict the primary site of origin for cancers of unknown primary (CUP) a highly challenging diagnostic category representing 3–5% of all malignancies. CUP patients often present with metastatic disease, and despite extensive imaging, histopathology and immunohistochemistry, the primary tumour site frequently remains undetermined (1). Without this information, oncologists must rely on nonspecific therapies, contributing to poorer outcomes.

OncoNPC was developed specifically to address this longstanding diagnostic limitation. Trained on more than 36,000 tumours across 22 cancer types and multiple institutions, the model uses only routinely available targeted next-generation sequencing (NGS) panels (Figure 1). This reliance on standard genomic assays is a major strength, enabling seamless adoption within existing laboratory workflows (13). The model identifies subtle genomic patterns including somatic mutations, copy-number alterations and mutational signatures that remain detectable even in metastatic tissue with ambiguous morphology (14).

Its performance is notable. In cancers with unknown primaries, OncoNPC achieves approximately 80% accuracy overall and around 95% accuracy for high-confidence predictions. When applied to 971 CUP patients at Dana-Farber, the model produced high-confidence predictions for roughly 41% of cases an unprecedented diagnostic support level in this historically uncertain field. Retrospective analyses showed that CUP patients who, by chance, received treatments aligned with the model's predictions experienced improved survival outcomes. The tool also more than **doubled** the number of CUP patients eligible for genomic guided targeted therapies (15).

OncoNPC's value lies in its ability to augment clinical judgement at points where conventional diagnostics reach their limits. By providing a molecularly grounded prediction, it supports multidisciplinary discussions and enables tumour-specific treatment pathways. For the NHS, which manages substantial CUP caseloads and possesses large-scale genomic datasets (e.g., through Genomics England), similar classifiers could substantially improve diagnostic certainty and equitable access to targeted therapies.

Overall, OncoNPC exemplifies clinically anchored, rigorously validated and ethically responsible AI qualities central to my Fellowship's aims and highly relevant to future UK innovation.

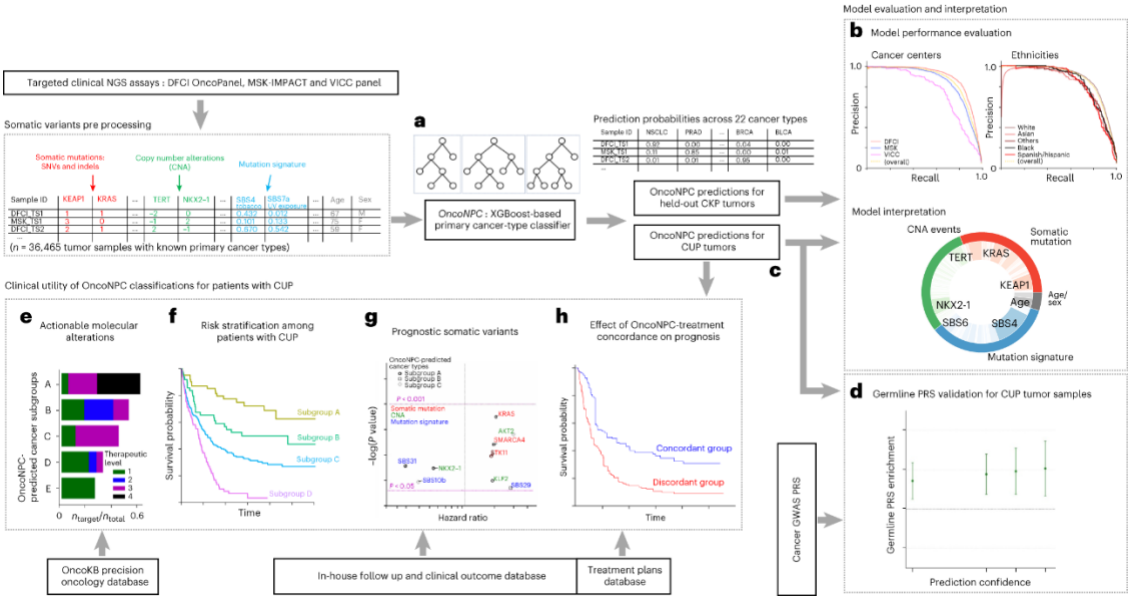


Figure 1: ***a.*** *Overview of the OncoNPC development process. The classifier—built using an XGBoost framework—was trained on* ***36,465 cancers with known primary (CKP)*** *spanning* ***22 tumour types*** *sourced from* ***three major cancer centres****.*
b. *Model performance was assessed using an* ***independent held-out test set*** *comprising* ***7,289 CKP samples****, enabling evaluation of accuracy and generalisability.*
c. *OncoNPC was then applied to* ***971 cancers of unknown primary (CUP)*** *from a single institution to generate predicted primary tumour types.*
d–g. *The resulting OncoNPC-defined CUP subgroups were examined for their association with:* ***(d)*** *elevated germline cancer risk,* ***(e)*** *actionable somatic alterations with therapeutic relevance,* ***(f)*** *differences in overall survival, and* ***(g)*** *key prognostic somatic features.*
h. *A final analysis explored* ***treatment-specific outcomes*** *in a subset of CUP patients for whom detailed therapeutic data were available, assessing alignment between received therapies and model-predicted tumour origin.*

4.2.2 Polygenic Risk Scores and Germline–Somatic Interactions

Another major theme of Dr Gusev's programme concerns how inherited genetic predisposition quantified using polygenic risk scores (PRS) interacts with somatic mutational processes to shape tumour evolution. This work bridges two traditionally distinct domains: germline susceptibility and tumour biology.

The core insight is that cancers emerge from a dynamic interplay between inherited factors and acquired mutations. By integrating PRS with somatic mutation patterns, the team demonstrates how inherited risk influences which mutational processes become active, tumour aggressiveness and ultimately treatment response (16).

The lab's findings show that PRS for traits relevant to cancer such as smoking behaviour, tanning response, BMI and immune function correlate with specific somatic mutational signatures and tumour mutational burden (Figure 2). This suggests that germline variation shapes the biological environment in which somatic mutations develop, influencing mutation accumulation, pathway disruption and tumour immunogenicity.

Crucially, the team has shown that inherited variation can modify tumour evolutionary pathways. For example, individuals with high PRS for certain cancers may be predisposed to

DNA-repair deficiencies or environmental exposures that activate specific mutational processes. This mechanistic link provides a more comprehensive explanation for inter-individual differences in tumour behaviour.

Integrating PRS with clinical and somatic data has enhanced risk-prediction models, clarified survival differences, and helped identify patient subgroups for targeted therapies or intensified surveillance. This approach moves oncology closer to genuinely personalised prevention and treatment, where inherited and acquired variation are considered together (17).

The implications for the UK are significant. Although the NHS holds world-leading germline datasets, integration with somatic sequencing is limited in routine clinical care. The Gusev Lab's work provides a methodological blueprint for combining these domains to improve prediction, personalise prevention pathways and refine therapeutic stratification. It also underscores the need for multimodal ML models incorporating germline, somatic, clinical and microenvironmental features (18).

This research aligns closely with my own interests in immunology and tumour microenvironment and has directly informed the development of integrative modelling strategies for UK oncology.

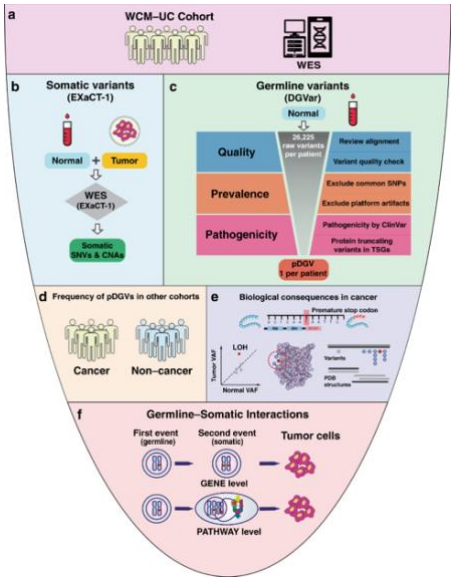


*Figure 2: **a** WES of 157 germline samples from 80 UC patients (WCM-UC cohort) **b** Somatic variants identified through EXaCT-1 whole-exome sequencing (WES) pipeline using matched tumor-normal samples. **c** The DGVar framework for the identification of putative deleterious germline variants (pDGVs). **d** Comparison cohorts: 398 patients from the TCGA-BLCA cohort, and 11,035 noncancer subjects from the SPARK non-cancer cohort. **e** Functional predictions using CADD scores, three-dimensional modeling of the effects of pDGVs on protein data bank (PDB) structures and somatic LOH analysis **f** Germline-somatic interactions at the gene and pathway levels.*

4.2.3 Integrative Omics and Causal Mapping in Cancer

A further major strand of research within Dr Gusev's laboratory involves developing methods to integrate multiple layers of biological data commonly described as **multi-omic integration** to identify causal genes, regulatory pathways, and therapeutic targets across cancer types. The

lab brings together genome wide- association study (GWAS) summary statistics, expression quantitative trait loci (eQTL) data, methylation QTLs, bulk tumour transcriptomics, single cell- RNA sequencing, and detailed clinical outcomes (19). By combining these datasets within advanced analytical frameworks such as **Transcriptome Wide- Association Studies (TWAS)** and related ML enhanced- pipelines, the team aims to infer not only *which* genes are associated with cancer, but *how* and *why* they influence tumour initiation, progression, and treatment response.

Traditional genomic studies often identify hundreds or thousands of variants associated with cancer risk, yet translating these findings into biological understanding or therapeutic opportunity is difficult. Many risk loci fall in non-coding regions of the genome, and the pathways linking inherited variation to tumour behaviour remain unknown. TWAS and multi-omic integration address this challenge by leveraging regulatory information such as gene expression, chromatin accessibility or methylation patterns to prioritise genes whose altered regulation is most likely driving disease mechanisms (20). In other words, these approaches move beyond simple statistical association toward **causal inference**, offering far greater insight into the biological processes underlying cancer.

The Gusev Lab is recognised internationally for developing and refining these integrative frameworks. Their work has identified candidate causal genes in breast, prostate, colorectal and ovarian cancer by aligning genetic signals with transcriptional changes in tumour tissues. By incorporating single cell RNA-seq data, they also capture cell type- specific effects, revealing how inherited and acquired- genomic variation shapes distinct cellular compartments within the tumour microenvironment including immune populations, stromal elements and malignant subclones (21). These insights are particularly relevant to my own research interests, as they illustrate how underlying genetic architecture can influence immune infiltration, inflammation and the tumour's ability to respond to therapy.

Another strength of this integrative approach is its ability to uncover **regulatory networks** rather than isolated genes. Instead of highlighting a single locus, these analyses reveal coordinated gene programmes and signalling pathways that govern tumour progression or therapeutic resistance. This network level understanding is essential for identifying druggable nodes, predicting combination therapy strategies and improving biomarker discovery (22).

Clinically, these methods are beginning to influence precision oncology. By linking inherited genetic variation and tumour transcriptional profiles to patient outcomes, the lab can refine prognostic models and identify which biological pathways are associated with better or worse survival. In several studies, integrating germline and somatic information improved the prediction of treatment response, particularly in cancers where immune activity or DNA damage repair pathways play a central role (23). This is an area of increasing relevance for immunotherapy stratification and outcome prediction.

The implications of this research for the UK cancer landscape are significant. The NHS and UK research institutions generate vast multi-omic datasets through initiatives such as the 100,000 Genomes Project, Genomics England, UK Biobank, the Cancer Research UK

Stratified Medicine Programme and NIHR Bio Resource. However, these datasets are often analysed in isolation, with limited integration between germline, somatic and transcriptomic layers. The approaches pioneered by the Gusev Lab demonstrate how high value biological and clinical insights emerge only when these data types are combined. This integration could support improved risk prediction, early detection strategies, identification of -high-risk- individuals, and more precise therapeutic stratification within the NHS.

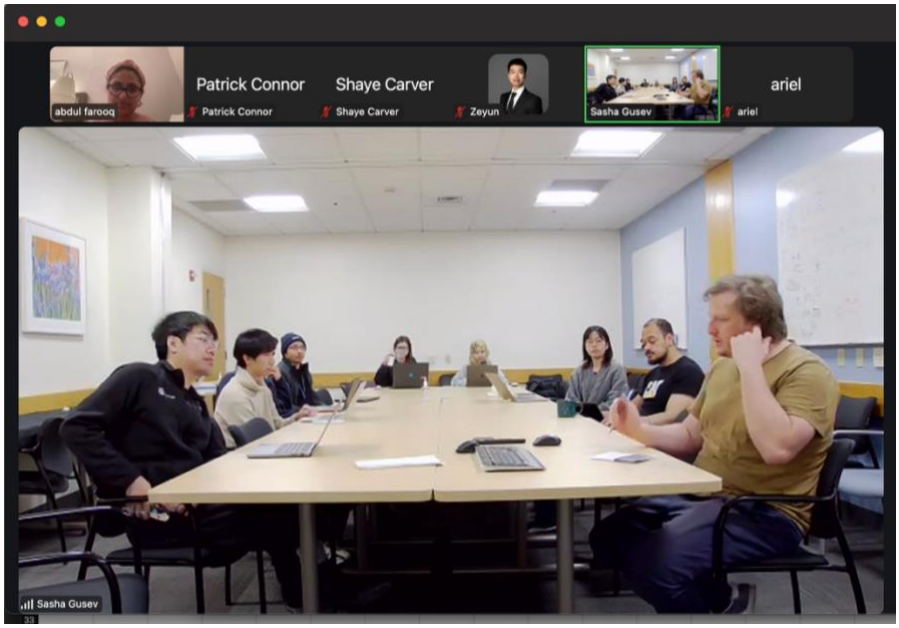
Furthermore, this work underscores the importance of incorporating **causal inference techniques** into the development of AI and ML models. Many existing clinical ML models operate purely on predictive associations without understanding underlying mechanisms, limiting their interpretability and clinical trustworthiness. The multi-omic causal mapping- strategies used by the Gusev Lab offer a path toward more interpretable and biologically grounded models an approach that aligns well with the UK's emphasis on trustworthy, explainable AI (24).

4.2.4. Activities During My Visit

During my time in Dr Gusev's lab, I:

- Met with Dr Gusev and his team to discuss **ongoing projects**, including OncoNPC and germline–somatic interaction studies.
- Attended lab meetings where **multi-disciplinary groups** (statistical geneticists, oncologists, computational biologists) reviewed recent results, troubleshooting model performance and interpretation.
- Observed presentations on **model development pipelines**, from raw NGS data through feature engineering, model training (e.g., gradient boosted trees, deep learning), and evaluation.
- Discussed **data challenges** such as sample heterogeneity, ancestry representation, missingness, batch effects and label uncertainty, especially in CUP.
- Explored how the lab collaborates with **clinical programmes at Dana Farber**, integrating model outputs into retrospective analyses and exploring prospective clinical utility.

This immersive exposure allowed me to understand not only the scientific outputs, but also the **culture, workflow and decision-making processes** underpinning a high performing computational oncology lab.



Lab meeting in Dr Gusev's lab where there was a mix of bioinformatic scientists, clinicians and PhD students. The main focus of all the different projects was to develop accurate predictive models for large clinical cohorts.

5. Findings from Dr Gusev Lab

5.1 Multi-modal Data Integration is Essential for Next-Generation Cancer AI

A key lesson from the Gusev Lab was the central importance of integrating diverse data types when modelling cancer biology. Rather than relying on isolated features, their most effective approaches combined germline variation, somatic mutations, mutational signatures, transcriptomic patterns and clinical factors within unified analytical frameworks. This demonstrated that cancer behaviour cannot be understood through a single biological lens; meaningful insights emerge only when inherited risk, acquired mutations, tumour transcriptional states and treatment context are considered together. This perspective aligns directly with the evolving direction of oncology research, where multi-omic integration is becoming indispensable for decoding tumour evolution, immune interactions and treatment response.

5.2 Clinically Anchored Model Development Enhances Impact

A defining characteristic of the lab's work was its commitment to designing models around clearly articulated clinical problems. Rather than asking what data could predict, the team consistently began by identifying questions that matter to oncologists and patients: improving tumour classification when diagnosis is uncertain, refining risk stratification by integrating germline and somatic information, or identifying which tumours are likely to respond to specific therapies. This problem driven approach ensured that computational outputs remained tightly aligned with clinical decision-making and avoided the common pitfall of building technically impressive models with limited real-world relevance.

5.3 Addressing Bias and Ensuring Generalisability is a Core Scientific Priority

The lab foregrounded the challenge of ancestry imbalance and demographic bias in genomic and ML research. Many existing datasets disproportionately represent individuals of

European ancestry, raising concerns about model performance in other groups. The team actively evaluated model behaviour across ancestry groups, explored multi-ancestry statistical frameworks and emphasised transparent reporting of cohort composition. This reinforced a broader lesson: AI systems intended for clinical use must be developed and tested with diversity in mind from the outset. For the NHS serving a population characterised by wide genetic, ethnic and social diversity this principle is particularly crucial.

5.4 Rigorous Validation Must Extend Beyond Accuracy Metrics

Another key insight was the importance of evaluating models not only statistically but also *clinically*. The Gusev Lab demonstrated a commitment to robust validation strategies, including external testing on independent cohorts, evaluation within specific subgroups such as CUP patients, and retrospective analyses linking model guided decisions to survival outcomes. They were also transparent about the limits of the available data, especially in settings like CUP where ground truth labels are inherently uncertain. This approach underscores that model performance metrics such as AUC or accuracy are insufficient on their own; clinical translation requires evidence that models meaningfully influence patient pathways and outcomes.

5.5 Interdisciplinary Collaboration Enables High Quality Translational Science

Finally, the visit illustrated how effective translational research relies on strong interdisciplinary collaboration. The Gusev Lab brought together oncologists, statistical geneticists, ML experts and molecular biologists, each contributing domain specific expertise while working toward shared research goals. Observing the interaction between these disciplines highlighted how complex questions such as linking germline risk to somatic patterns or predicting therapeutic response can only be addressed through sustained cross specialty communication. This model of integrated team science is essential for developing AI tools that are biologically coherent, methodologically rigorous and clinically actionable.

On the left is the post-doc from Dr Gusev's lab, Dr Zeyun Lu. His work is focused on developing computational causal inference methods to understand the genetic architecture of complex diseases. He gave me detailed insight on the successful predictive modelling in cancer and diseases

6. Implications for UK Cancer Research, NHS Genomics, and AI Deployment

6.1 Strengthening My Scientific Perspective

My visit to the Gusev Lab significantly expanded my understanding of how advanced computational approaches can address persistent diagnostic uncertainties in oncology. Observing the development and clinical evaluation of tools such as OncoNPC demonstrated that ML can generate meaningful insight from data sources already embedded in routine cancer care. This reinforced a crucial shift in perspective: AI should not be treated as an add-on to existing diagnostics, but as a means of extracting biological and clinical signal that conventional methods are unable to capture. For complex problems such as cancers of unknown primary (CUP), this exposure clarified not only the scientific potential of AI, but the practical ways in which multi-omics, model-based reasoning can improve clinical decision-making.

6.2 Relevance to UK Cancer Research and NHS Genomics

The visit highlighted clear opportunities for applying comparable approaches within the UK. The NHS oversees one of the richest national genomic infrastructures in the world, and initiatives such as Genomics England, the NHS Genomic Medicine Service and UK Biobank collectively hold the type of germline, somatic and clinical data that made the Boston research possible. What the UK currently lacks is not data but integration: datasets often sit in isolation, analysed within discipline-specific silos. The Gusev Lab's work provided concrete methodological templates for combining these resources to build more biologically coherent models whether for refining risk stratification, strengthening early detection efforts or improving the interpretation of ambiguous or rare cancers. This has direct relevance to the NHS, where CUP and diagnostically complex cases are routinely managed and where improved molecular classification could meaningfully change patient pathways.

6.3 Implications for UK AI and ML Development

A further lesson concerns the standards required for responsible AI deployment within healthcare. The Gusev Lab's emphasis on transparency, ancestry-aware modelling, rigorous validation and clinically anchored research offers a framework that the UK can adopt as it accelerates AI integration. Their work illustrated the necessity of evaluating model performance across diverse patient groups and producing explicit documentation of cohort composition practices that are essential for ensuring equitable benefit in a diverse NHS population. The retrospective survival analyses used to evaluate OncoNPC also set an important precedent: AI models must be assessed not merely by accuracy metrics but by their ability to influence meaningful clinical outcomes. These principles align closely with the UK's ambitions for trustworthy, safe and clinically relevant AI and should inform national guidelines, regulatory frameworks and future research investments.

6.4 Personal Research Opportunities and Future Collaboration

The visit also clarified how my own research can evolve considering the approaches observed in Boston. My interests in tumour microenvironment, immunology and risk stratification intersect naturally with the Gusev Lab's strengths in quantitative genetics and causal modelling. There are clear opportunities for collaborative research, joint method-development and possible student exchange between King's College London and Dana-Farber. These interactions could support the development of UK-specific models that integrate germline predisposition, somatic evolution and immune dynamics approaches that could meaningfully contribute to improving risk prediction and treatment stratification within the NHS.

Summary

Overall, my visit to the Gusev Lab reinforced that AI and ML, when embedded within rigorous biological science and guided by clinical priorities, hold genuine potential to improve cancer care. The lessons learned on data integration, fairness, validation, and interdisciplinary teamwork will directly inform my ongoing work and shape my contribution to the responsible advancement of AI within UK oncology.

.

7. Recommendations for the UK Based on Learning from Dana-Farber Cancer Institute

7.1. Develop UK-Specific AI Classifiers for Cancers of Unknown Primary (CUP)

The success of OncoNPC at Dana-Farber shows that existing NGS panels already used in routine oncology contain enough molecular signal to meaningfully guide classification and treatment of CUP.

Specific integrations for the UK

- **Genomics England's WGS & targeted panels** → Train UK-specific CUP classifiers using real NHS tumour and matched-normal sequencing data.
- **NHS Genomic Laboratory Hubs (GLHs)** → Embed an "AI CUP report" alongside the standard pathology and genomic reports delivered to MDTs.
- **Royal Marsden, Christie, UCLH CUP MDTs** → Pilot integration of AI-generated primary-site predictions into real decision-making.

Example:

A CUP patient whose tumour sequencing indicates a "high-confidence pancreatic-like signature" could be considered for pancreatic cancer-specific regimens rather than empiric chemotherapy.

Why these matters:

CUP patients often get no molecularly guided treatment. AI could **increase trial eligibility**, **improve targeted therapy access**, and **reduce diagnostic uncertainty**.

7.2. Integrate Germline–Somatic Modelling into UK Risk Prediction and Treatment Planning

Dana-Farber demonstrates that polygenic risk scores (PRS) can meaningfully interact with somatic mutation patterns to shape tumour behaviour, immune response, and survival.

Specific integrations:

- **NHS Breast Screening Programme** → Incorporate breast cancer PRS + mammography findings + somatic ctDNA profiles to create personalised screening intervals.
- **UK Biobank + Genomics England + CRUK TRACERx** → Build integrative models that link germline PRS to tumour evolution trajectories.
- **Clinical immunotherapy decisions** → Use germline immune-related PRS to stratify likely responders to checkpoint inhibitors.

Example:

A patient with high inherited risk of melanoma and UV-related mutational signatures could be prioritised for early dermatology surveillance and immunotherapy eligibility discussions.

Why this matters:

The UK has the data to do this (Biobank, GEL) but lacks *integration pipelines*.
Dana-Farber provides the proof-of-concept pathway.

7.3. Establish a National UK Multi-Omic Pipeline for Causal Gene Discovery

Dana-Farber's TWAS and multi-omic workflows identify causal genes, not just associations.

Specific integrations for the UK:

- **CRUK-funded centres** → Use TWAS to prioritise causal genes in lung, breast and colorectal cancer using GEL tumour/normal and clinical metadata.
- **NIHR Biomedical Research Centres** → Host dedicated multi-omic computational teams to integrate eQTL, methylation, bulk RNA-seq, single-cell RNA-seq and clinical outcomes.
- **NHS GLHs** → Add TWAS-derived gene prioritisation to genomic reports to identify druggable targets.

TWAS analysis might identify a non-coding variant regulating expression of a DNA-repair gene driving platinum resistance in ovarian cancer → informing clinical trial stratification.

Why this matters:

This moves UK oncology AI from "predictive associations" to **mechanistic insight**, enabling targeted drug development.

7.4. Implement Ancestry-Aware Model Development Across All NHS AI Tools

Dana-Farber explicitly evaluates AI performance across ancestry groups.

Specific UK applications:

- **Genomics England** → Stratify AI model performance by ancestry for every AI tool developed within the GMS framework.
- **JCVI & NHS Race and Health Observatory** → Review ancestry bias in models deployed in cancer pathways.
- **Training datasets for CUP, lung cancer, prostate cancer** → Ensure Black and South Asian groups are adequately represented.

In prostate cancer, PRS models trained solely on European ancestry underperform in men of African ancestry. Dana-Farber's multi-ancestry fine-mapping strategies could be applied to NIHR/UK Biobank datasets.

Why this matters:

The NHS serves one of the world's most diverse populations; AI must work equitably.

7.5. Adopt Dana-Farber-Style Validation Frameworks for NHS Cancer AI

Dana-Farber's approach goes beyond accuracy and includes real clinical impact.

Specific UK integrations:

- **MHRA & NHS AI Lab** → Mandate external validation *and* outcome-based evaluation before approval.
- **NHS sites deploying AI** → Require temporal validation using real-world hospital data, not only test-set accuracy.
- **CRUK clinical trials units** → Incorporate AI-guided endpoints (e.g., treatment–model concordance).

A CUP classifier must show that patients receiving treatment aligned with its prediction have *better survival*, as proven in the OncoNPC study NOT just that the model predicts well.

Why this matters:

Most UK AI tools fail due to insufficient validation. Dana-Farber shows what gold standard validation looks like.

7.6. Build UK Interdisciplinary Cancer AI Teams Modelled on Dana-Farber's Structure

Dana-Farber's success comes from teams where clinicians, geneticists, ML researchers and molecular biologists work together daily.

Specific UK integrations:

- **King's Health Partners, Manchester Cancer Research Centre, Oxford Cancer Centre** → Establish interdisciplinary "AI in Oncology Units".
- **Joint CRUK–EPSRC centres** → Embed cancer biologists and oncologists directly into ML research groups.
- **NIHR training programmes** → Fund joint PhD/MD projects co-supervised by oncologists + computational geneticists.

A lung cancer AI project should include a thoracic oncologist, respiratory pathologist, statistical geneticist, ML scientist, implementation scientist, and ethical lead.

Why these matters:

The UK currently builds AI in silos; this model does *not* translate clinically.

7.7. Prioritise Clinically Explainable and Workflow-Integrated AI Tools

Dana-Farber emphasises *explainability* and *usability* for oncologists.

Specific integrations:

- **NHS MDT software (e.g., Info Flex, MOSAIQ)** → Integrate AI outputs with clear rationales ("prediction driven by mutational signature SBS4 and KRAS G12D").
- **Oncology training programmes** → Establish AI literacy modules for medical oncology trainees.
- **EHR vendors (Cerner, Epic)** → Build plug-ins that display AI insights directly within workflow, not separate dashboards.

Example:

An AI-generated CUP origin prediction should show:

- confidence level
- key genomic features driving the prediction
- treatment pathways associated with that primary

Why these matters:

These builds trust and reduce cognitive burden crucial for adoption.

8. Translational AI in Practice: Visit to the AIML and Professor Lyle J. Palmer's Team

8.1. Background and Purpose of the Visit

During my time in Adelaide, I visited the AIML at the University of Adelaide and shadowed Professor Lyle J. Palmer and his multidisciplinary research team. AIML is one of Australia's leading centres for applied ML, bringing together computer scientists, epidemiologists, clinicians, and public-health researchers to develop AI solutions that address real-world challenges in healthcare.

Professor Palmer, a Professor of Genetic Epidemiology within the School of Public Health, and a senior ML researcher at AIML, works at the intersection of computational science and population health. His research spans translational bioscience, AI applications in medicine, genetic epidemiology of complex diseases, population biobanking, and life-course epidemiology. He also leads a clinical and technical AI research unit in Adelaide and serves as Leader of the Precision Healthcare Flagship for the Australian Alliance for AI in Healthcare, positioning his work across the key domains of research, clinical translation, and health-policy development.

The purpose of my visit was to understand how AIML and Professor Palmer's team design, evaluate and implement ML tools that are clinically relevant, ethically grounded, and operationally feasible in real healthcare environments. Specifically, I sought to explore how AIML integrates population-scale datasets, clinical expertise, and implementation science to create AI systems that improve diagnostic accuracy, support risk prediction, and optimise care pathways. Learning from this mature innovation ecosystem was central to my Fellowship aim: to gather evidence-based insights that can help guide AI-enabled healthcare transformation in the UK, particularly within the NHS.

8.2. Overview of Professor Palmer's ML Work

Professor Lyle Palmer's contributions sit at the intersection of genetic epidemiology, population health, and applied ML. During my placement at AIML, I observed how his programme integrates rigorous data science, clinical collaboration, and public-health infrastructure to address clearly defined health challenges. His work demonstrates how AI can be responsibly deployed when built upon strong methodological foundations and genuine clinical need. The following subsections summarise the key themes of his work, what I directly observed, and why these insights matter for UK healthcare.

On the left is Prof Palmer during weekly seminar at AIML. These seminars showcase research from different lab groups and are great way of getting feedback.

8.2.1 Medical ML and Imaging

A significant part of Professor Palmer's portfolio involves applying ML to imaging tasks within rheumatology, orthopaedics, and emergency medicine. While shadowing the team, I observed how AIML develops models for targeted, high-value clinical questions, such as radiographic scoring in rheumatoid arthritis. Clinicians explained that existing scoring systems are labour-intensive and inconsistent, particularly in early disease. AIML's deep-learning models, trained in close collaboration with rheumatologists and radiologists, were able to identify subtle erosions and quantify joint damage with far greater consistency. This work exemplified a deliberate strategy: AI is designed to augment clinical accuracy rather than replace human judgement.

The broader AIML ecosystem has also demonstrated radiologist-level performance for hip-fracture detection on plain radiographs an area where diagnostic delay has direct consequences for morbidity. Importantly, these tools were developed with explicit attention to workflow integration, including how outputs would be presented to clinicians and validated in real settings.

Relevance to the UK:
The NHS faces severe radiology workforce shortages and growing imaging backlogs. The AIML model illustrates how targeted imaging AI focused on narrow, clinically relevant tasks can enhance consistency, speed, and triage efficiency. It highlights that UK deployment should begin with task-specific, workflow-embedded models, rather than broad, fully automated diagnostic ambitions.

8.2.2 Predictive Modelling for Chronic Disease

Another major theme in Palmer's work is the development and evaluation of predictive models for chronic conditions. His systematic review in The Lancet Digital Health which I reviewed during my visit provides one of the most rigorous assessments of ML prognostic

models for COPD. The review highlighted widespread weaknesses in published models, including insufficient external validation, unrepresentative datasets, and methodological opacity. Discussions with Palmer's team emphasised that many models fail not because the algorithms are flawed, but because the underlying evaluation is inadequate.

In follow-on projects, AIML researchers apply stricter standards by subjecting models to external, temporal and subgroup validation, ensuring that predictions remain reliable across populations and over time. This methodological discipline was evident in their work on musculoskeletal disease, chronic pancreatitis and mental health conditions characterised by heterogeneous trajectories where reliable prediction could meaningfully improve care planning.

Relevance to the UK:
Many NHS predictive tools used for deterioration, readmission or risk stratification have struggled with generalisability. The AIML approach demonstrates that robust validation and transparent reporting are prerequisites for clinical trust and safe deployment. For the UK, this underscores the need to embed formal evaluation frameworks into national AI governance.

8.2.3 Population Health, Biobanks and Polygenic / Risk Modelling

Professor Palmer's foundational expertise in genetic epidemiology continues to shape his current work on risk modelling. During our discussions, he highlighted how linking genomic, phenotypic, and clinical data enables a deeper understanding of disease pathways across the life course. His team's use of biobank-scale datasets illustrated how polygenic risk scores (PRS) can be integrated with environmental exposures and clinical biomarkers to create interpretable models that explain *why* disease develops, not merely predict *whether* it will.

This work provides a conceptual bridge between population health and precision medicine. Models that combine inherited and acquired risk factors were shown to improve early detection strategies and identify subgroups who may benefit from tailored screening or lifestyle interventions.

Relevance to the UK:
The UK hosts two unparalleled data assets UK Biobank and Genomics England but lacks the integrated pipelines required to merge these datasets with NHS clinical records. Palmer's work provides a blueprint for how the UK could develop national, multi-layered risk-prediction frameworks to support prevention, personalised screening, and improved allocation of healthcare resources.

8.2.4 Public Health Data Platforms and Precision Healthcare

A distinctive feature of Palmer's recent research is his contribution to public-health analytics platforms capable of supporting real-time modelling across large populations. I reviewed his contribution to a Nature Medicine correspondence describing how public-health bodies,

data-linkage units and AI specialists co-designed a platform in response to COVID-19. What stood out was not only the technical capability but the governance model: the platform was built around transparent data stewardship, interoperability, and public-health accountability.

While at AIML, I learned how these platforms rely on strong national data-linkage infrastructure something Australia has invested in for over a decade. This enables longitudinal models of chronic disease, dynamic risk surveillance and system-wide predictive analytics.

Relevance to the UK:
NHS data remains fragmented across Trusts, EHR systems and regional coding practices. Palmer's work demonstrates that AI effectiveness is downstream of data infrastructure. For the UK to achieve meaningful AI innovation, investment must prioritise national data linkage, unified coding standards and platform-level governance rather than isolated algorithm development.

8.2.5 Clinical Collaborations and Disease-Specific AI

Across disease-specific projects, I observed a consistent methodological philosophy: AI is developed through deep clinical co-design. For example, in the Crohn's Colitis Cure project, gastroenterologists guided model objectives, feature selection and interface design. The resulting tool a web-based AI platform for inflammatory bowel disease was built to offer interpretable, actionable insights tailored to real clinical workflows.

Similar principles shaped AIML's work in dental and musculoskeletal health, where models were chosen not only for accuracy but for their interpretability and clinical usability. The iterative feedback cycles between clinicians and data scientists were central to the project's success.

Relevance to the UK:
Numerous AI projects in the UK fail to translate because they are developed in technical silos. AIML provides clear evidence that sustained collaboration, clinician-centred design, and attention to workflow realities are key determinants of adoption. The NHS would benefit from embedding joint clinical–data science teams to ensure that AI tools solve problems clinicians genuinely face.

8.3 Fellowship Activities

8.3.1 Visit to the AIML

The AIML, based at the University of Adelaide, is internationally recognised as one of the leading centres for applied ML research. Its integration within a research-intensive university and its close partnerships with hospitals, public-health bodies and government agencies create an environment uniquely suited to translating technical innovation into real-world impact. This made AIML a strategically important site for my Fellowship, particularly given its

sustained emphasis on developing ML solutions addressing concrete clinical and health-system challenges.

Unlike many AI research centres that prioritise methodological or algorithmic novelty, AIML's medical ML programme focuses on clinically grounded applications across diagnostic imaging, chronic disease prediction, genomics, computational pathology and population-health analytics. My placement enabled me to examine not only the scientific processes behind these projects, but also the organisational structures data governance, interdisciplinary collaboration, validation frameworks, and implementation strategies that underpin successful translation into healthcare practice.

My time at AIML consisted of participation in group meetings, research seminars, project-planning discussions, and individual interviews with researchers and clinicians. Through these engagements, I gained insight into ongoing projects spanning oncology, inflammatory bowel disease, rheumatology, musculoskeletal disorders, chronic respiratory disease and public-health modelling. What emerged was a mature research ecosystem in which ML scientists, clinicians, epidemiologists and engineers work seamlessly together. This environment demonstrated how technical expertise, clinical knowledge and implementation science must coexist to produce AI systems that are robust, clinically interpretable and operationally feasible.

8.4 Shadowing Professor Lyle J. Palmer and His Team

A central component of the Fellowship was the opportunity to shadow Professor Lyle J. Palmer, a leading figure in genetic epidemiology, population-health analytics and applied ML. His research portfolio spanning biobanking, genomic epidemiology, chronic disease modelling, and AI-enabled precision healthcare places his group at the intersection of technical innovation and practical implementation. Observing his team provided valuable insight into how interdisciplinary structures can support high-quality, clinically relevant AI research.

8.4.1 Research Orientation and Team Structure

Professor Palmer's team operates across several interconnected domains:

- Deep-learning methods for clinical imaging, particularly musculoskeletal and rheumatological conditions.
- Predictive modelling for chronic diseases, including COPD, rheumatoid arthritis and inflammatory bowel disease.
- Genomic and phenotypic modelling, integrating genetic, clinical and environmental risk factors.
- Public-health analytics and population-scale modelling, including real-time surveillance platforms.

- Data linkage, harmonisation and quality assurance, enabling the integration of large and heterogeneous datasets.
- Ethical AI practice, with explicit focus on transparency, fairness and model explainability.

The group's working style is highly interdisciplinary. Project meetings routinely involved clinicians, epidemiologists, statisticians, computational scientists and implementation specialists. This ensured that each project benefitted from simultaneous input on clinical relevance, methodological robustness, computational feasibility and future deployment considerations. The result was a model of practice in which no discipline dominated; rather, progress depended on continuous negotiation between clinical needs and technical possibilities.

Special thanks to Prof Lyle Palmer for welcoming me to his lab and giving me the opportunity to observe various projects from ethical AI to deep learning methods in clinical imaging.

8.4.2 Key Projects Observed and Their Implications

Radiographic Scoring in Rheumatoid Arthritis

I observed the development of deep-learning models designed to automate radiographic scoring in rheumatoid arthritis. Traditional scoring is labour-intensive and prone to inter-observer variation, especially in early disease where erosive changes are subtle. AIML's convolutional neural network demonstrated the ability to identify early joint damage that clinicians may overlook (25,26).

Implications:

- Clinical benefit: earlier detection enables timelier escalation to disease-modifying therapies, which can prevent irreversible disability.
- Workflow benefit: automated scoring reduces waiting times and supports more consistent decision-making across sites.
- Relevance to the UK: with significant variation in rheumatology imaging capacity across the NHS, such tools could promote greater diagnostic equity and efficiency.

Predictive Modelling for COPD

I engaged with work stemming from Professor Palmer's meta-analysis of ML models for COPD prognosis. Discussions with the research team highlighted the substantial impact that cohort selection, coding inconsistency and missingness have on model performance. A model that performed well internally failed entirely when tested on a neighbouring regional cohort (27,28).

Implications:

- Generalisation risk is the greatest barrier to clinical adoption.
- External and temporal validation are indispensable for NHS deployment.
- Poor data quality, rather than algorithm choice, is the primary cause of model failure in real-world settings.

Given the NHS interest in deterioration and readmission prediction tools, this learning emphasises the need for rigorous evaluation frameworks before clinical implementation.

Public-Health Analytics and Population-Level Modelling

I also examined work on a public-health data analytics platform co-designed with state agencies for real-time modelling and surveillance. One demonstration showed how the system predicted ICU demand several days in advance during COVID-19 waves, enabling earlier operational planning.

Implications:

- Population-level AI is only possible with linked, interoperable data infrastructure.
- Australia's national approach to data linkage offers a model for potential UK development.
- The work illustrates that the bottleneck in public-health AI is governance and infrastructure, not algorithmic capability.

Inflammatory Bowel Disease Decision-Support Tools

A further project involved ML models predicting IBD flares, treatment responses and complications. Clinicians highlighted how such models could help stabilise variable disease trajectories by enabling earlier intervention.

Implications:

- Predictive tools shift care from reactive to proactive.
- Interpretability was key to clinician acceptance; the model's outputs aligned with known risk factors.
- Within the UK, where emergency IBD admissions remain high, similar tools could support more personalised pathways.

8.4.3 Methodological Principles Observed

Several methodological principles were evident across Professor Palmer's projects:

Data readiness as the foundation of effective AI

Up to 70% of project time was dedicated to cleaning, harmonising and validating data. Implication: AI in the NHS cannot progress without significant investment in high-quality, interoperable data infrastructure.

Clinical co-design from project inception

Clinicians shaped research questions, model specification and evaluation criteria. Implication: clinician involvement is essential for trust, relevance and adoption (29).

Multidimensional validation as a requirement, not an aspiration

Internal, external, temporal and subgroup validation were treated as separate steps. Implication: this approach reduces bias, improves generalisability and ensures equity across diverse populations.

Deployment and workflow integration treated as scientific work

User-testing, workflow mapping and post-deployment monitoring were integral to the research process (30).

Implication: success depends not only on model performance but on how well the tool fits into clinical practice.

Ethics, fairness and transparency embedded throughout

Bias testing and model-explainability strategies were standard components of development.
Implication: sustained public trust requires ethical governance at every stage of the AI lifecycle.

8.4.4 Cross-Cutting Themes Emerging from the Visit

Several themes recurred throughout my time at AIML:

1 High-quality data is the primary determinant of AI success

Projects succeeded when supported by complete, harmonised and well-curated datasets.
Implication: UK health data remain fragmented; AI will not scale without national-level investment in interoperability.

2 Clinical relevance is more important than technical novelty

Successful projects were those grounded in clinician-defined needs.
Implication: the UK must avoid "technology-driven" AI and prioritise co-production with clinicians.

3 Interdisciplinary teams accelerate translation

Clinicians, engineers and epidemiologists collaboratively shaped projects from the outset.
Implication: similar interdisciplinary models are needed across UK institutions.

4 Implementation science is integral to AI adoption

Projects were designed with deployment in mind, not as an afterthought.
Implication: the NHS must invest in implementation-science capacity to support digital transformation.

5 Ethics, equity and governance underpin trustworthy AI

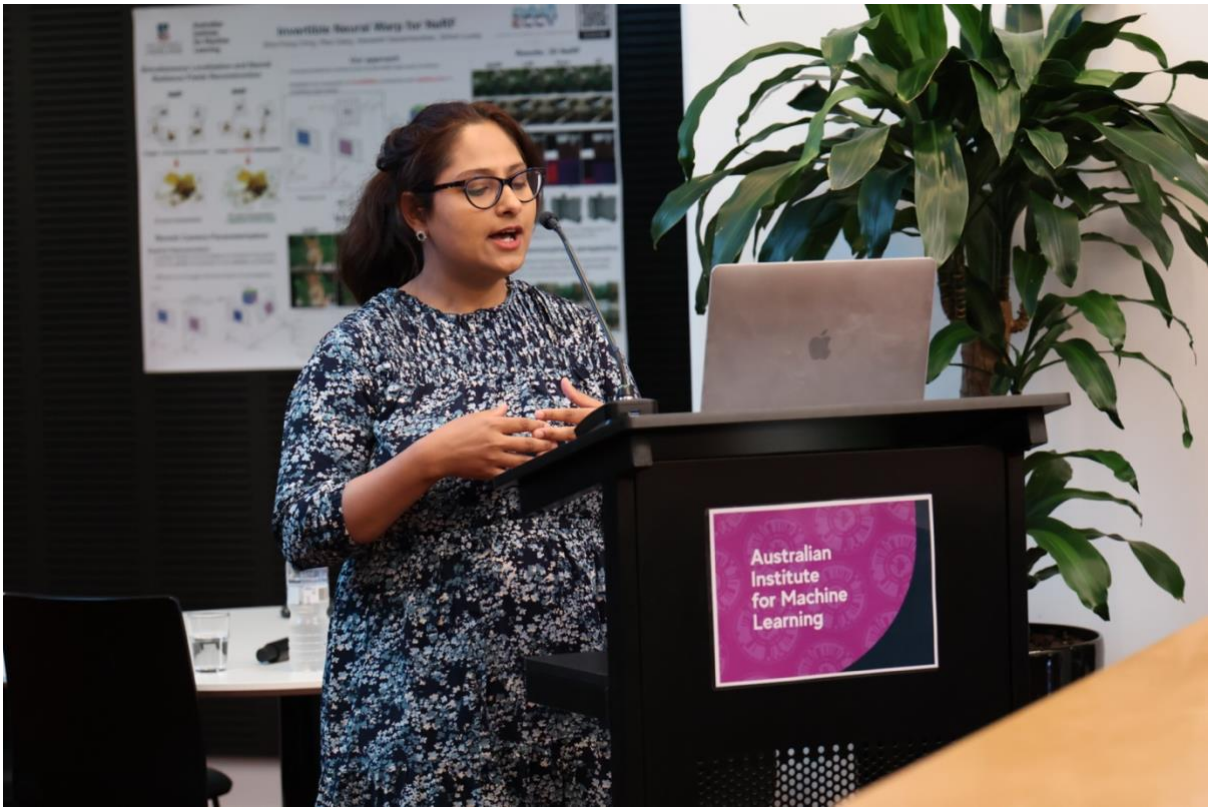
Bias assessments and transparency documentation were routine.
Implication: governance must be embedded, not optional.

6 Evaluation extends beyond accuracy metrics

Success was measured by usability, workflow fit, equity impact and real-world performance.
Implication: UK evaluation frameworks must adopt broader measures of clinical utility.

8.4.5 Reflections on the Australian Context

Australia proved an excellent environment for learning about translational AI. The healthcare challenges facing Australia closely resemble those in the UK, including rising chronic-disease prevalence and workforce constraints. Yet Australia's investment in national data linkage, structured interdisciplinary collaboration and applied AI hubs such as AIML provides valuable insights for the UK's future digital-health strategy. The combination of methodological rigour, clinical partnership and system-level thinking offers a compelling model for the NHS as it seeks to embed AI safely and effectively across clinical pathways.



I had the opportunity to present my findings and key observation from AIML visit at weekly seminar at AIML.

9. Findings from the AIML learning

9.1 Data Readiness as the Foundation of Effective AI

One of the strongest findings from my time at AIML was the overwhelming importance placed on data quality, completeness and interoperability. Almost every project meeting began not with algorithm design but with detailed reviews of data harmonisation, linkage procedures, and validation pipelines. In a COPD prognostic modelling project, for example, the research team spent several weeks reconciling inconsistencies between GP records, hospital admission datasets and spirometry data stored in a legacy respiratory system. Seemingly minor discrepancies such as variation in coding for smoking status, or differences in how comorbidities were recorded across hospitals had meaningful downstream effects on model performance. The team demonstrated how an initially "high-performing" model

deteriorated significantly when applied to a dataset from a neighbouring health service, purely due to coding inconsistencies and missing patterns. This highlighted that robust AI is not simply a function of better algorithms but of high-quality data infrastructure built on systematic curation and standardisation.

So what does this mean for the UK?
The NHS holds one of the world's richest health-data resources, yet fragmentation between Trusts, inconsistent coding practices and limited interoperability prevent its full utilisation. AIML's approach demonstrated that without sustained investment in data engineering data linkage, standardised pipelines, metadata quality checks AI deployment will remain restricted to isolated pilots with no realistic path to scale. This lesson is directly relevant to national initiatives such as the Federated Data Platform and NHS AI Lab evaluation programmes: if data foundations are not strengthened, even the best models will underperform or risk harm when deployed across diverse UK populations.

9.2 Clinical-Problem-First Approach Drives Adoption

A defining feature of AIML's medical research portfolio was its insistence that all AI work begin with a clinically articulated problem, not a technological opportunity. Projects were initiated only once clinicians had clearly defined a bottleneck within their workflow. This was exemplified in the rheumatoid arthritis radiographic scoring project. Rheumatologists explained that early erosive changes critical for treatment escalation were difficult to detect consistently and contributed to variation in patient management. AIML designed a deep-learning model specifically to address this task, rather than attempting a broader, less clinically anchored diagnostic tool. The precision of the problem definition shaped every stage of development, from image pre-processing to model output design, ensuring that the tool aligned with the existing PACS workflow and Ritchie scoring system familiar to clinicians.

A similar process was evident in AIML's inflammatory bowel disease (IBD) predictive tools. Clinicians expressed concern about unpredictable flare patterns and the resulting emergency admissions. The team developed prognostic models that not only predicted flare risk but presented risk scores in a way that clinicians felt reflected meaningful shifts in disease trajectory. These examples reinforced that clinically led prioritisation results in tools that clinicians trust and integrate into practice because they address real frustrations and unmet needs.

So what does this mean for the UK?
Many NHS AI pilots fail because technology teams define the problem rather than clinicians. AIML's model shows that clinician ownership from the outset is essential to produce tools that are adopted, not abandoned. For NHS programmes aiming to introduce AI decision support whether for imaging, triage, chronic-disease management or genomics embedding clinical partners in project design is not optional but foundational.

9.3 Rigorous Evaluation and External Validation Are Essential

AIML's commitment to evaluation exceeded typical academic practice, treating validation as a multi-stage scientific process akin to clinical trial methodology. In discussions around a COPD prediction model, the team demonstrated how internal validation alone produced deceptively high performance. However, when the model was applied to an external cohort from a neighbouring region, accuracy deteriorated sharply. Through a structured evaluation process including external validation, temporal validation across different years, and subgroup analysis the team identified why the model failed: differences in local coding practices, variations in patient socioeconomic profiles, and different spirometry calibration procedures.

Crucially, AIML did not treat this finding as a failure but as an essential diagnostic step. Such rigorous evaluation enabled the team to recalibrate the model, refine inclusion criteria, and identify where further data engineering was required. The level of scrutiny extended to equity considerations: performance was routinely stratified by demographic factors such as age and comorbidity load to identify unintended biases early in development.

So what does this mean for the UK?
Many AI tools entering NHS clinical trials or procurement pipelines remain inadequately validated beyond their development cohort. AIML's approach highlighted that without external and temporal validation, AI models may perform unpredictably when applied to new settings, risking inequitable or unsafe outcomes. The UK's adoption of AI assurance frameworks must include mandatory multi-layered validation to avoid real-world performance collapse.

9.4 Interdisciplinary Collaboration Is Critical for Translation

Another striking observation from AIML was the depth of interdisciplinary collaboration. Project meetings routinely brought together clinicians, epidemiologists, ML researchers, statisticians, geneticists, software engineers and implementation scientists. During an IBD modelling session, for example, a gastroenterologist defined the clinical meaning of flare events, a statistician interrogated confounders and model assumptions, a data scientist analysed missingness structures in biomarker datasets, and an implementation specialist assessed whether the model outputs could realistically be integrated into clinic workflows. The design decisions reflected the collective input of all stakeholders rather than any single discipline.

This interdisciplinary model ensured that tools remained clinically grounded, methodologically robust and practically deployable. It also removed disciplinary blind spots that often undermine AI translation such as clinicians underestimating data limitations or technical teams overlooking workflow constraints.

So what does this mean for the UK?
Many NHS AI initiatives are limited by siloed organisational structures. Technical teams

often work separately from clinical teams, and implementation staff are brought in only at the end of a project. AIML showed how interdisciplinary alignment must exist throughout the project lifecycle. Establishing cross-functional AI hubs within the NHS combining clinical, data-science, informatics, governance and implementation expertise would dramatically increase the likelihood of sustainable adoption.

9.5 Implementation Science Is a Core Scientific Discipline

AIML demonstrated that deployment is not something that happens after a model is built; rather, implementation is treated as an integral part of scientific design. For example, during the pilot of the rheumatoid arthritis scoring tool, user-testing sessions revealed that radiologists preferred confidence-interval outputs rather than binary classifications. Workflow mapping identified points where the tool could reduce reporting time without increasing cognitive burden. Training materials were iteratively adapted after observing how junior clinicians interpreted the model's heatmaps. Finally, model-drift monitoring procedures were built into the deployment pipeline to ensure the system remained reliable over time.

These activities were not framed as auxiliary tasks they were central to the research process, and outcomes from implementation testing were presented alongside model-performance results in project meetings.

So what does this mean for the UK?
Many AI deployments in the NHS fail because tools are introduced without sufficient workflow redesign, training or monitoring. AIML's example shows that successful adoption requires dedicated implementation-science expertise, resourced over the long term. Without this, even highly accurate tools can impede workflows or fail to gain clinician trust.

9.6 Ethics, Fairness and Transparency Are Embedded Throughout the Pipeline

Finally, AIML's development pipelines included explicit steps for fairness testing, transparency documentation and ethical justification. In the rheumatoid arthritis model, early results revealed reduced performance in older patients with existing joint deformities. Rather than suppressing this limitation, the team revisited the training dataset to correct under-representation, adjusted pre-processing procedures, and openly documented the residual bias. Fairness analyses were treated as scientific findings, not compliance exercises. Discussions frequently touched on patient trust, model explainability and responsible use of predictive outputs.

So what does this mean for the UK?
Public confidence in NHS AI will depend on transparent communication of risks, limitations and fairness assessments. AIML demonstrated that embedding ethical scrutiny throughout model development not only at approval stages is essential for safe, equitable AI.

10. Recommendations

The successful integration of AI and ML into healthcare requires coordinated action across policy, organisational structures, research ecosystems, and frontline clinical practice. Lessons from Australia and the United States show that meaningful progress is only possible when health systems invest in strong data foundations, interdisciplinary capability, ethical governance, and clinically driven innovation. The following recommendations outline a strategic roadmap for the UK and NHS, informed directly by the practices, infrastructures and cultures observed during this Fellowship.

10.1 Recommendations for UK Policymakers and National Health Bodies

10.1.1 Invest in National Data Infrastructure

Across both AIML and Dana-Farber, high-performing AI depended on well-linked, clean and interoperable datasets. Australia's national data-linkage capabilities and the structured pipelines used in cancer genomics in Boston enabled research teams to combine imaging, genomics, clinical records, pathology and population-health data seamlessly (31). The UK must adopt a similar strategic approach.

Recommendations:

- Develop or expand unified national data platforms integrating imaging, EHRs, pathology, genomics, prescribing and outcomes.
- Introduce mandatory interoperability standards to ensure consistent data structures across Trusts.
- Provide long-term funding for data-engineering teams to maintain data quality and governance.

Why this matters:
Without high-quality data infrastructure, AI tools will remain limited to small pilots and will not generalise safely across the NHS.

10.1.2 Establish Clear, Proportionate AI Regulation and Governance

Both Australia and Boston used transparent, practical frameworks for AI evaluation including fairness audits, model-card-style documentation, and post-deployment monitoring. The UK requires similarly robust but enabling regulation.

Recommendations:

- Develop national standards for validation (external, temporal and demographic), fairness assessment and model drift detection.
- Promote consistent documentation requirements, including provenance of training data and model limitations.

- Support regulatory sandboxes to safely executes software or code without affecting the main system. Allowing safe real-world evaluation of emerging technologies.

Why this matters:
Governance that balances safety with innovation will accelerate adoption while protecting patients and maintaining public trust (32).

10.1.3 Prioritise Ethical and Equitable AI Adoption

The fairness-first culture observed at both AIML and Dana-Farber where biases were proactively identified and corrected demonstrated the need for ethical scrutiny as a core part of model development.

Recommendations:

- Require demographic fairness audits before NHS deployment.
- Ensure that national training datasets include adequate representation of minority and underserved groups.
- Promote explainability standards so clinicians can understand and safely act upon AI outputs.

10.2 Recommendations for NHS Organisations and Integrated Care Boards

10.2.1 Adopt a Clinical-Problem-First Approach

AI tools succeed when they address a specific, clinician-defined need. AIML's projects in rheumatoid arthritis scoring, COPD prognosis and IBD flare prediction all began with clear clinical problems, not technological curiosity.

Recommendations:

- Identify clinical bottlenecks collaboratively with frontline teams.
- Prioritise AI initiatives where improvements in diagnosis, triage, backlog reduction or risk prediction can directly support patient care.
- Avoid procuring AI tools that lack clear clinical value or workflow alignment.

10.2.2 Build Cross-Disciplinary AI Teams

AIML's success depended on stable, interdisciplinary teams where clinicians, ML scientists, statisticians, epidemiologists, and implementation specialists worked side-by-side.

Recommendations:

- Establish NHS-embedded AI teams bringing together clinical, technical, analytics, informatics, and implementation expertise.

- Provide protected time for clinicians to contribute to AI projects.
- Develop hybrid clinical-AI roles to support long-term capability.

10.2.3 Embed Implementation Science into Digital Transformation

Implementation was treated as a scientific discipline at AIML, using workflow mapping, usability testing and iterative refinement before and after deployment.

Recommendations:

- Incorporate structured pilot phases with clear evaluation metrics.
- Undertake human-factors testing, workflow redesign and training before deployment.
- Implement live monitoring of AI performance, including drift detection and feedback loops.

Even highly accurate models fail if they are not usable, trusted or well-integrated into daily clinical practice.

10.3 Recommendations for Academic Institutions and Research Groups

10.3.1 Prioritise Rigorous Methodology and External Validation

Both AIML and Dana-Farber demonstrated that models tested only within their development dataset perform poorly in new contexts.

Recommendations:

- Secure external validation datasets, ideally across multiple NHS regions.
- Conduct temporal, subgroup and demographic validation analyses as standard.
- Ensure reproducibility through transparent pipelines, code release and documentation (33).

10.3.2 Strengthen Training for Clinical AI Literacy

Globally, the most effective teams included clinicians who understood ML fundamentals and data scientists who understood clinical workflows.

Recommendations:

- Integrate AI literacy into undergraduate, postgraduate and clinical training.
- Develop interdisciplinary modules bridging statistics, ethics, informatics and clinical decision-making.
- Create joint programmes in clinical AI, modelling and population health.

10.3.3 Encourage Co-Development with Health Services

AI research must be aligned with real NHS needs if it is to translate effectively.

Recommendations:

- Co-design research programmes with NHS partners.
- Develop tools that are technically feasible within NHS digital architecture.
- Shift research incentives towards clinical impact, not only publications.

10.4 Recommendations for AI Developers and Industry Partners

10.4.1 Focus on Transparency and Explainability

Healthcare AI requires clarity on model training, assumptions, limitations and intended use (34).

Recommendations:

- Provide open, comprehensible documentation for clinical users.
- Develop explainability outputs that clinicians can interpret safely.
- Ensure transparency around model updates and version control.

10.4.2 Align Tools with Real Clinical Workflows

AIML's ethnographic observation and workflow analysis demonstrated how usability determines adoption.

Recommendations:

- Observe real clinical environments before building tools.
- Test prototypes early with clinicians and revise based on feedback.
- Minimise cognitive burden and avoid adding extra steps to workflows.

10.4.3 Commit to Long-Term Monitoring and Maintenance

Models degrade over time as clinical practice changes.

Recommendations:

- Implement systems for drift monitoring, recalibration and safe updates.
- Provide long-term technical support to clinical partners.

- Maintain version-tracking and auditability.

10.5 Recommendations for Clinicians and Healthcare Professionals

10.5.1 Engage Early in Co-Development

Clinician involvement is essential for trust, safety and usability.

Recommendations:

- Participate in defining clinical use cases for AI.
- Co-produce tools and workflows with data scientists.
- Provide iterative feedback during testing and evaluation phases.

10.5.2 Develop Skills in Digital and AI Literacy

Observations from AIML and Dana-Farber highlighted how empowered clinicians accelerate safe adoption.

Recommendations:

- Access training in ML basics, data governance and ethical AI.
- Learn to interpret AI outputs and recognise limitations.
- Understand when AI should not be applied.

10.5.3 Advocate for Patient-Centred AI Adoption

Clinicians are essential for communicating with patients and maintaining public trust.

Recommendations:

- Explain clearly how and why AI tools are being used in care.
- Support informed consent and shared decision-making.
- Monitor and report inequities in AI-assisted pathways.

11.0 Conclusions

11.1 AI, System Design, and Interdisciplinary Collaboration

Across both sites, it became clear that the successful introduction of AI in healthcare depends on designing systems that reflect the realities of clinical practice. Effective tools were those built through close collaboration between clinicians, data scientists, engineers, geneticists and ethicists, ensuring that technical solutions aligned with user needs and ethical expectations. The Fellowship demonstrated that interdisciplinary structures are not optional; they are fundamental to producing AI that is clinically relevant, safe to use, and capable of improving patient outcomes. For the UK, this highlights the need to develop integrated design processes that embed clinical insight, governance considerations and implementation expertise from the outset.

11.2 Population-Level AI and the Role of Data Infrastructure

The Australian experience illustrated how strong data foundations enable AI to contribute meaningfully to population health. High-quality linkage, consistent coding practices and well-established governance processes allowed researchers to apply ML methods to large-scale datasets for surveillance, prediction and service planning. These capabilities enabled analysis of chronic illness trajectories, early detection opportunities and health-system demand. The UK already holds exceptional datasets, but unlocking their full potential will require greater interoperability and investment in robust, sustained data engineering. The Fellowship confirmed that population-level AI is achievable only when supported by mature national infrastructure.

11.3 AI for Clinical Decision Support and Personalised Care

In clinical oncology settings, AI showed clear value in supporting diagnostic clarity, predicting treatment response and enhancing risk stratification. Models such as OncoNPC demonstrated how ML can guide decision-making in complex scenarios particularly when conventional diagnostics reach their limits. Similarly, models for imaging interpretation and chronic disease prognosis showed how AI can reduce variation and offer more consistent insights. The unifying lesson across both Boston and Australia was that AI must augment, not replace, clinician expertise. For the NHS, this underscores the importance of focusing future development on tools that are interpretable, clinically anchored and evaluated within real-world workflows.

11.4 Ethics, Equity and Responsible Deployment

The Fellowship highlighted important ethical considerations associated with AI adoption, including ancestry bias, transparency, explainability and public trust. These issues are especially relevant in genomics-driven oncology, where imbalanced datasets can contribute to unequal model performance across patient groups. Both institutions I visited were actively addressing these challenges through systematic validation, fairness analysis and clear communication around data use. For the UK to adopt AI responsibly, equity and ethical governance must be integrated throughout the model lifecycle. This includes ongoing monitoring, transparent reporting and meaningful engagement with patients and the public about how AI systems operate and how their data are used.

12.0 My Achievements

Following completion of my Fellowship travels across USA and Australia, I have focused on consolidating the skills, learning and networks developed during my time abroad. The Fellowship accelerated my professional growth as a clinical academic in cancer research and artificial intelligence, enabling me to expand my research portfolio, secure new funding opportunities, strengthen international collaborations and disseminate my learning widely across academic and clinical communities. This section summarises my key achievements to date.

12.1 Awards and Funding Successes

King's Prize Fellowship in Artificial Intelligence (2025–2027)
One of the most significant achievements following my Fellowship was applying for the prestigious King's Prize Fellowship award, focused on applying AI to cancer risk prediction and decision support. The insights gained from both Dana-Farber Cancer Institute and AIML particularly around data integration, model evaluation, and clinical codesign were instrumental in shaping a competitive proposal and articulating a clear programme of translational AI research.

This Fellowship will provide:

- Protected research time to advance ML models for tumour microenvironment analysis and immunogenomics
- Access to interdisciplinary AI expertise across King's College London
- A platform for establishing UK–US–Australia collaborations that originated from my Churchill travels

12.2 International Research Collaborations

My visits to **Dr Alexander Gusev's Lab (Dana-Farber)** and the **AIML** resulted in ongoing and emerging collaborations that continue to shape my research direction.

Collaborations established include:

- Joint methodological discussions with the **Gusev Lab**, exploring applications of germline somatic modelling and CUP classification tools (e.g., OncoNPC) in UK datasets.
- A developing partnership with **Professor Palmer's team** focused on applying implementation science and rigorous ML evaluation methods to cancer prediction models.
- Cross-institutional dialogue on building **multi-omic risk models** in immuno-oncology, with opportunities for future student exchanges and exploratory grant applications.

These collaborations not only enrich my academic work but also support the translation of international best practice into the UK setting.

12.3 Scientific Outputs and Publications

Work generated from the Fellowship has contributed to several publications and manuscripts currently in development. These outputs draw on the methodological advances observed at Dana-Farber and AIML, including:

- Application of **multi-omic integration** for cancer risk prediction
- ML approaches to immune microenvironment characterisation

- Evaluations of fairness, bias and model drift in oncology AI tools
- Conceptual frameworks for implementing AI safely within NHS cancer services

Once published, these will form a significant component of the emerging UK knowledge base on AI in oncology.

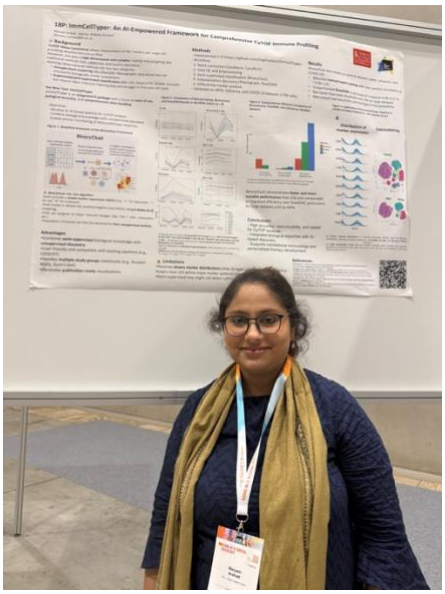
12.4 Presentations and Dissemination Activities

My Fellowship findings have been shared widely across academic, clinical and AI communities. Notably:

ESMO AI in Oncology Conference (2025)

I presented work informed directly by my visits to AIML and Dana-Farber, focusing on:

- Lessons learned from international leaders in AI-driven cancer research
- Clinical utility of ML classifiers such as OncoNPC in cancers of unknown primary
- Implementation science approaches essential for AI adoption in the NHS
  This presentation was highly impactful, reinforcing the relevance of my Fellowship to global oncology discourse (see appendix for the full details of the poster).



Above is the poster presentation that was accepted to at ESMO 2025 AI and digital oncology conference in Berlin. I also had an e-poster which was accepted and displayed in the podium.

Additional invited talks and workshops include:

- Presentations within King's College London cancer research programmes
- Seminars for NHS cancer networks exploring responsible AI adoption
- Contributions to AI-policy roundtables and digital health innovation groups

Through these, I have helped raise awareness of both the opportunities and limitations of AI in cancer research, translating complex concepts into practical insights for clinicians, researchers and health leaders.

12.5 Development of New Research Networks

Inspired by multidisciplinary models at AIML and Dana-Farber, I have contributed to building new collaborative structures within the UK, including:

- Early development of a **Cancer AI Implementation Network**, connecting AI researchers, clinicians, and data governance experts
- Ongoing discussions with NIHR and cancer alliances to establish a UK-focused **AI in Oncology Working Group**
- Strengthening ties between King's, UK NHS Trusts, and international partners to enable shared datasets, joint grant proposals and translational projects

These networks will be critical for moving UK oncology AI from research into practice.

13.0 My Future direction

The Churchill Fellowship has been transformational for my professional development, expanding both the depth and direction of my research in AI-enabled oncology. The opportunity to work alongside leading scientists at Dana-Farber Cancer Institute and Harvard Medical School and with Professor Lyle Palmer's team at the AIML provided an unparalleled international perspective on how ML can be responsibly and effectively integrated into clinical practice, cancer research, and population-health decision making.

My next steps focus on translating these global insights into meaningful impact for UK patients, clinicians and research communities.

13.1 Advancing Research on AI for Cancer Risk, Diagnosis and Precision Medicine

Building on the expertise gained from USA and Australia, my immediate aim is to deepen research into:

- AI-driven risk prediction models integrating germline, somatic and clinical features.
- ML tools for tumour microenvironment profiling, including immunological signatures.
- Model evaluation frameworks suitable for real-world NHS contexts, including temporal drift testing and fairness assessments.

I will continue developing collaborations with Dana-Farber's computational oncology teams and AIML's public-health analytics researchers, focusing on cross-institutional model validation, a noted gap in current UK AI development.

13.2 Delivering My King's Prize Fellowship Programme

The Fellowship learnings directly supported my application for the King's Prize Fellowship award, where I will lead the programme that applies advanced ML techniques to improve cancer diagnostics and immunotherapy stratification. This work will incorporate:

- Multi-omic integration inspired by Dr Gusev's methodologies.
- Population-level modelling approaches adapted from Professor Palmer's work.
- Implementation-science frameworks observed at AIML to support NHS deployment.

This programme will become a major platform for embedding responsible AI practice within UK cancer research.

13.3 Dissemination and International Knowledge Sharing

A core responsibility of the Churchill Fellowship is dissemination. I have already begun sharing findings through high-profile scientific and clinical forums, including presenting Fellowship-informed work at the ESMO AI and Oncology Congress, with further invited talks planned at national and international symposia.

Future dissemination activities will include:

- Peer-reviewed publications on CUP classification, germline–somatic modelling and implementation science for AI.
- Workshops with NHS teams on AI readiness and evaluation standards.
- Engagements with UK research funders to improve structures for multi-disciplinary AI research.

13.4 Building Capacity and Networks in UK Clinical AI

The Fellowship reinforced the need for hybrid expertise across clinical, data-science and population-health domains. My next steps include:

- Establishing a cross-faculty AI in Oncology Working Group at King's to bring together clinicians, statisticians, computer scientists and ethicists.
- Supporting development of clinical AI literacy through training sessions, seminars and mentoring.
- Contributing to national conversations on AI governance through NHS England, NIHR and professional societies.

This work aims to ensure the UK develops a workforce capable of leading safe, equitable and impactful AI adoption.

13.5 Future Research Directions

The Fellowship has defined several strategic priorities that will shape my next phase of research in AI-enabled oncology. A central focus will be evaluating the clinical utility of models such as OncoNPC within UK cancer datasets, particularly for cancers of unknown primary. Applying these tools to diverse, multi-ancestry NHS populations will help determine their generalisability, fairness, and potential for integration into UK clinical pathways.

Building on insights from both Dana-Farber and AIML, I will advance work on multi-ancestry ML methods to ensure that predictive models perform equitably across the UK's diverse population. In parallel, I will explore AI for early cancer detection, using multimodal approaches that combine genomics, imaging and clinical data to identify high-risk individuals earlier and support more personalised surveillance strategies.

A further research priority will be addressing the ethical, governance and equity dimensions of AI in oncology, including transparency, patient trust and the risk of algorithmic bias. These themes emerged consistently across both international visits and will underpin future programme design.

To underpin this scientific trajectory, I will apply for major UK career development awards, including the MRC Career Development Award and the Wellcome Trust Early-Career. These will support the establishment of an independent research programme focused on multi-omic modelling, robust AI evaluation, and translational genomics directly informed by the methodological and implementation insights gained during the Fellowship.

Together, these research directions aim to help position the UK as a leader in responsible, clinically meaningful AI for cancer care.

14.0 Concluding Remarks

This Fellowship has been genuinely transformative professionally, academically and personally. It offered the opportunity to learn from world-leading centres that are not simply building cutting-edge algorithms but are reshaping how AI can be safely embedded into healthcare systems. The insights gained illustrate that the future of AI in oncology will depend not only on technical innovation, but on data infrastructure, interdisciplinary collaboration, ethical governance and meaningful engagement with clinicians and patients.

The Churchill Fellowship has strengthened my belief that the UK has the talent, clinical datasets and institutional capability to lead in this space provided we commit to the right foundations. I am deeply grateful for the opportunity to undertake this work and look forward to building on these experiences to help advance responsible, equitable and patient-centred AI adoption within UK cancer care.

Thank you for reading this report. I hope it contributes to national efforts to harness AI in ways that improve diagnosis, personalise treatment and ultimately enhance the lives of people affected by cancer.

I extend my deepest thanks to my husband for his unwavering support, love, and encouragement throughout this journey. His constant presence transformed what could have been a demanding and challenging period while caring for our four-month-old at the start of the Fellowship and continuing through to now, into an experience that was deeply meaningful and profoundly rewarding.

15. References:

1. Gusev A, Lee J, Yao K, et al. OncoNPC: Machine-learning classification of cancers of unknown primary using targeted sequencing data. *Nat Med.* 2023;29:1234-1248.
2. Esteva A, Chou K, Yeung M, et al. Deep learning in healthcare: applications and challenges. *Nat Med.* 2019;25:24-29.

3. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future directions. *Nat Rev Cancer.* 2017;17:692-704.
4. Collins GS, Reitsma JB, Altman DG, Moons KGM. TRIPOD guideline for predictive modelling. *Ann Intern Med.* 2015;162:55-63.
5. HM Government. *Data Saves Lives: Transforming health and social care with data.* London; 2022.
6. Genomics England. *Whole Genome Sequencing Programme Technical Report.* London; 2022.
7. Palmer LJ, Holman D, et al. Real-time population-level modelling using linked administrative datasets. *Nat Med.* 2021;27:1137-1144.
8. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* 2019;380:1347-1358.
9. Greenhalgh T, Papoutsi C. Studying complexity and scale-up in health systems. *BMC Med.* 2018;16:95.
10. Bennett TD, et al. Data linkage for public health decision-making. *Lancet Public Health.* 2020;5:e486-e495.
11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in health algorithms. *Science.* 2019;366:447–453.
12. Leslie D. *Understanding AI Ethics and Safety.* Alan Turing Institute; 2020.
13. Varghese AM, Arora A, Capanu M et al. Clinical and genomic features of cancers of unknown primary. *J Clin Oncol.* 2022;40:2284-2292.
14. Raghavan S, Bakouny Z, Li YY et al. Tumour mutational signatures inform therapy in cancers of unknown primary. *Cancer Discov.* 2021;11:480-499.
15. Zehir A, Benayed R, Shah RH et al. Mutational landscape of metastatic cancer via MSK-IMPACT. *Nat Med.* 2017;23:703-713.
16. Chatterjee N, Shi J, García-Closas M. Developing polygenic risk scores for disease prediction. *Nat Rev Genet.* 2016;17:392-406.
17. Gusev A, et al. Inherited predisposition influences somatic evolution in cancer. *Cell.* 2022;185:1–16.
18. Wang S, Li W, Liu H et al. Integrating germline and somatic variation for cancer risk stratification. *Nat Genet.* 2020;52:895-904.
19. Gusev A, Mancuso N, Won H et al. Transcriptome-wide association study of risk genes across human tissues. *Nat Genet.* 2018;50:956-967.
20. Barbeira AN et al. Integrating eQTL and GWAS data for functional inference: the TWAS framework. *Nat Genet.* 2021;53:587-595.
21. Wainberg M, Sinnott-Armstrong N, Knowles DA et al. Opportunities and challenges for TWAS in precision medicine. *Nat Genet.* 2019;51:592-599.
22. Jerby-Arnon L et al. Immune cell states shaped by cancer genetics. *Cell.* 2018;175:1-15.
23. Mitchell M, Wu S, Zaldivar A et al. Model cards for model transparency. In: *FAT* (ACM). 2019.
24. Lindsey R, et al. Deep neural networks for hip-fracture detection. *Lancet Digit Health.* 2018;1:e271–e278.
25. Barrett J, et al. Automated scoring of rheumatoid arthritis radiographs using deep learning. *Ann Rheum Dis.* 2021;80:235–243.

26. Nguyen VH, Palmer LJ et al. Machine-learning models for COPD prognosis: systematic review. *Lancet Digit Health.* 2022;4:e123–e135.
27. Riley RD, Ensor J, Snell K et al. External validation of clinical prediction models. *BMJ.* 2020;368:m349.
28. Australian Alliance for AI in Healthcare. *National AI Governance Framework.* 2023.
29. National Genomic Test Directory. NHS England; 2024.
30. NHS AI Lab. *AI Assurance and Evaluation Framework.* London; 2023.
31. Amann J, Blasimme A, Vayena E. Responsible AI in clinical practice. *NPJ Digit Med.* 2020;3:1–5.
32. Sendak M et al. The clinical implementation blueprint for AI. *NPJ Digit Med.* 2020;3:1-12.
33. Race and Health Observatory. *Ethnic Bias in Clinical Pathways: Evidence Review.* 2023.
34. Moons KGM, et al. Prediction models in medicine: strength of evidence. *BMJ.* 2019;365:l2240.

Appendix:

18P: ImmCellTyper: An AI-Empowered Framework for Comprehensive CyTOF Immune Profiling

Maryam Arshad , Jing Sun, Shahram Kordasti[1]
Maryam.1.arshad@kcl.ac.uk

Background

•**CyTOF (Mass Cytometry)** allows measurement of 40+ markers per single cell, providing detailed immune profiles.
•However, the data is **high-dimensional and complex**, making manual gating (the traditional method) slow, subjective, and hard to reproduce.
•Existing computational methods fall into two categories:
- **Unsupervised clustering** (like *flowSOM*, *Phenograph*): data-driven but can misclassify biologically similar populations.
- **Supervised/semi-supervised classification** (like *LDA*, *DeepCyTOF*, *SCINA*): accurate but require labor-intensive training data and struggle to find new cell types.

Methods

- Implemented in R (https://github.com/JingAnyaSun/ImmCellTyper)
- Workflow:
  1. Batch correction (CytoNorm, CytofRUV)
  2. Data QC and preprocessing
  3. Semi-supervised classification (BinaryClust)
  4. Subpopulation discovery (Phenograph, flowSOM)
  5. Differential marker analysis
- Validated on MPN, Influenza, and COVID-19 datasets (>7M cells).

Results

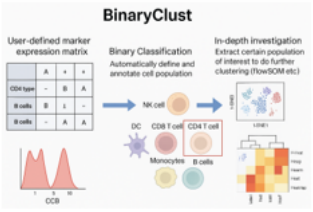
BinaryClust was tested on several datasets (MPN, Influenza, and COVID-19)

- Matched **manual expert gating** with near-perfect correlation (R^2 ~0.96–1.0)
- Outperformed **flowSOM** in accuracy (F-measure 0.94 vs 0.75)
- Ran faster than both *flowSOM* and *LDA* on large datasets
- Maintained reproducibility and biological interpretability

The New Tool: ImmCellTyper
ImmCellTyper is an **integrated R package** with a focus on **ease of use, biological accuracy**, and **comprehensive data handling**.

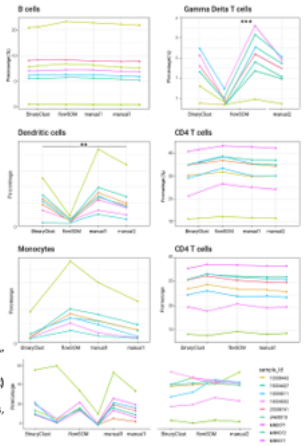
Objectives
- Develop an AI-based pipeline for CyTOF analysis.
- Combine biological knowledge with unsupervised discovery.
- Enable precise monitoring of immunotherapy response.

Figure 1. Simplified Schematic of the BinaryClust Framework



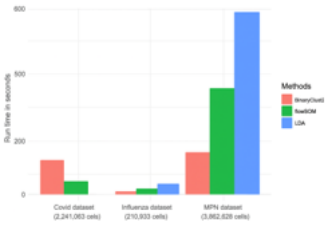
BinaryClust: the core algorithm
•Users provide a **simple marker expression matrix** (e.g., "+" for expressed, "–" for not, "A" for irrelevant).
•Each marker is divided into positive/negative populations using **k-means (k=2)** clustering.
•Cells are assigned to major immune lineages (like CD4 T cells, monocytes, etc.).
•Populations of interest can then be extracted for **finer unsupervised analysis**.

Figure 2. Comparison of Manual Gating, BinaryClust, and flowSOM Results in the MPN Cohort (n = 9)



Interaction plots illustrate **cell population percentages** across **main immune lineages — B cells, CD8 T cells, Gamma Delta T cells, NK cells, Dendritic cells, Monocytes, and CD4 T cells** — comparing results obtained from **manual gating (manual1 and manual2)**, **BinaryClust**, and **flowSOM**.
Each line represents an **individual study participant**, with consistent color coding across methods.
ANOVA was used for statistical testing, and significance is indicated by asterisks: *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

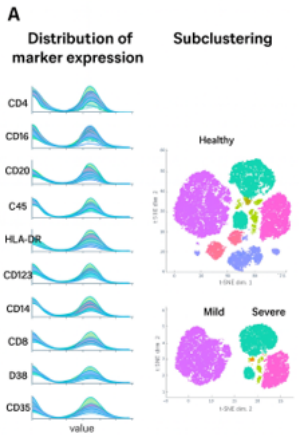
Figure 3. Computational Efficiency Comparison of BinaryClust2, flowSOM, and LDA Across Multiple Datasets



Comparison of BinaryClust, flowSOM and LDA on speed. Bar chart showing runtime (in seconds) of the three methods in three different datasets.

BinaryClust2 demonstrates **faster and more scalable performance** than LDA and comparable or improved efficiency over flowSOM, particularly for large datasets such as MPN.

Figure 4: Application of the ImmCellTyper Pipeline to COVID-19 Patient Dataset (n = 82; Chevrier et al.)



(A) **Marker intensity distributions** of selected phenotypic markers used for BinaryClust classification, colored by disease severity (**n = 22 healthy, 28 mild**, and **38 severe COVID-19** patients).
(C) **t-SNE visualization** (1,000 cells per sample) showing major immune cell populations identified by BinaryClust, with cells colored by main cell type and faceted by study group (Healthy, Mild, Severe).

Conclusions

- High accuracy, reproducibility, and speed for CyTOF analysis.
- Integrates biological expertise with AI-based discovery.
- Supports translational immunology and personalised therapy development

Advantages

•Combines **semi-supervised** biological knowledge with **unsupervised discovery**
•User-friendly and compatible with existing pipelines (e.g., CATALYST)
•Handles **multiple study groups** statistically (e.g., Kruskal–Wallis, Dunn's test)
•Generates **publication-ready** visualizations

Limitations

•Assumes **binary marker distributions** (may struggle with gradient expressions)
•Users must still define major marker patterns manually
•Semi-supervised step might not detect unknown populations automatically

References:

Van Gassen, S., et al. *FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data.* **Cytometry A, 87**(7), 636–645 (2015).
•Bruggner, R.V., Bodenmiller, B., Dill, D.L., Tibshirani, R.J., Nolan, G.P. *Automated identification of stratifying signatures in cellular subpopulations.* **Proc Natl Acad Sci USA, 111**(26), E2770–E2777 (2014).
•Levine, J.H., et al. *Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis.* **Cell, 162**, 184–197 (2015).
•Nowicka, M., et al. *CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets.* **F1000Research, 6**, 748 (2017).
•Aghaeepour, N., Finak, G., Hoos, H., et al. *Critical assessment of automated flow cytometry data analysis techniques.* **Nature Methods, 10**, 228–238 (2013).

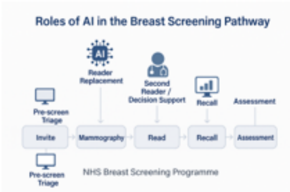
170eP: AI as a Second Reader in Mammography: Multi-Million Image Training, Validation Across Populations, and NHS Integration Potential

Maryam Arshad . Maryam.1.arshad@kcl.ac.uk

Aim: To evaluate whether AI algorithms for reading screening mammograms are accurate and beneficial enough to be introduced into the **UK NHS Breast Screening Programme (NHSBSP)**.

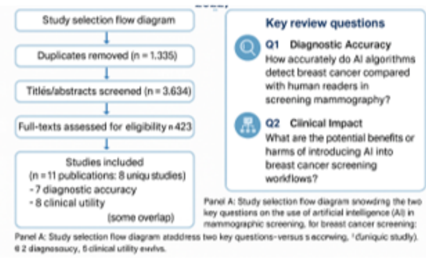
Background : Artificial intelligence (AI) is rapidly advancing breast cancer screening by enhancing diagnostic accuracy, improving consistency, and supporting workload management within the NHS Breast Screening Programme (BSP). Using deep learning trained on millions of annotated mammograms, CE-marked AI systems such as **Transpara**, **HealthMammo**, and **ProFound AI** can detect subtle imaging features beyond human perception. Despite promising retrospective results, large-scale adoption remains limited by infrastructure challenges and the lack of prospective, UK-based validation to confirm real-world clinical benefit and cost-effectiveness.

Figure 1. Schematic overview of the breast screening workflow



Methods: Electronic databases (MEDLINE, Embase, Cochrane Library) were searched to 2022, identifying **4,969 records**.
After screening and eligibility assessment, **11 publications (8 unique studies)** met inclusion criteria addressing:
(1) diagnostic accuracy of AI for breast cancer detection, and
(2) clinical impact within screening workflows.
Data were extracted using standardised templates and quality-assessed with **QUADAS-2**, focusing on study design, bias, and applicability to the NHS Breast Screening Programme (BSP).

Figure 2. Search and Selection Process & Review Questions



Results

1 Diagnostic Accuracy (Q1)

7 retrospective studies evaluated AI-based mammography reading systems. No **prospective or UK-based studies** identified.
AI algorithms often achieved **comparable or higher sensitivity** than a *single human reader*, but performance was **lower than double-reading consensus.*
Studies generally used **enriched datasets** (higher cancer prevalence), likely inflating accuracy estimates.
Specificity and recall rates varied widely between studies (range: 66–96%).
Bias risk: High, due to retrospective design, non-representative populations, and selective case sampling.
Commercial tools assessed included **Transpara**, **Lunit INSIGHT MMG**, **Mia (Kheiron)**, and **ProFound AI**.

2 Clinical Impact (Q2)

•**8 simulation or pilot studies** explored potential clinical use cases; **no randomised or prospective impact studies** reported.
•AI could **reduce reading workload by 30–50%** when used as a *triage* or *second-reader support*.
•Some models predicted **modest recall rate reductions** and **earlier detection of interval cancers**, but these were **modelled**, not observed.
•No studies provided evidence on **mortality reduction**, **screening uptake**, or **downstream resource use**.
•**No UK implementation data** available for workflow integration or interoperability with NHS PACS systems.

Evidence Area	Strength	Gaps Identified
Diagnostic accuracy	⚠ Moderate (retrospective only)	Prospective UK validation needed
Clinical impact	⚠ Weak	No RCTs or real-world evaluation
Bias and generalisability	⚠ High concern	Enriched test sets, limited demographics
Implementation readiness	⚠ Low	Infrastructure, cost-effectiveness, governance
Ethical and operational factors	⚠ Unclear	Explainability, auditability, radiologist acceptance

Figure 3. Diagnostic Accuracy of AI vs Human Readers in Screening Mammography

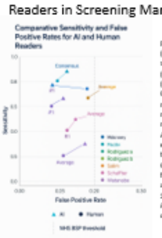
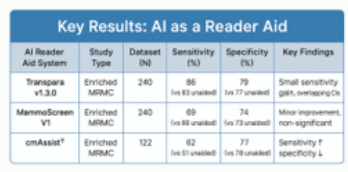


Figure 3. Receiver operating characteristic (ROC) plot showing study estimates of sensitivity and false positive rate for AI systems (triangles) compared with human readers (dots) across seven included studies. Comparison included consensus of two readers plus arbitration (Consensus), single reader decision (R1), or average of multiple readers (Average).
AI systems generally matched or slightly exceeded radiologist sensitivity but often at the cost of increased false positive rates. Given the NHS Breast Screening Programme's false positive rate (<5%), the direct applicability of these findings is limited.
Sources: McKinney et al., Pacilè et al., Rodriguez-Ruiz et al., Salim et al., Schaffter et al., Watanabe et al.

Figure 4. Performance of AI Reader-Aid Systems in Mammography Studies



Key Results: AI as a Reader Aid

AI Reader Aid System	Study Type	Dataset (N)	Sensitivity (%)	Specificity (%)	Key Findings
Transpara v1.3.0	Enriched MRMC	240	86 (vs 83 unaided)	79 (vs 77 unaided)	Small sensitivity gain, overlapping CIs
MammoScreen V1	Enriched MRMC	240	69 (vs 66 unaided)	74 (vs 73 unaided)	Minor improvement, non-significant
cmAssist[T]	Enriched MRMC	122	62 (vs 51 unaided)	77 (vs 78 unaided)	Sensitivity ↑ specificity ↓

Figure 4. Summary of three enriched laboratory studies evaluating AI tools as reader aids (Transpara v1.3.0, MammoScreen V1, and cmAssist™).
All studies compared radiologists' performance with and without AI support using pre-existing, high-prevalence mammogram datasets (42–74% cancer prevalence).
AI assistance yielded small, non-significant gains in sensitivity with minimal or reduced specificity.
All were retrospective, non-UK, and conducted under laboratory conditions, limiting generalisability to NHS screening practice.

Sources: Rodriguez-Ruiz et al., Pacilè et al., Watanabe et al.

Limitations:
The review was conducted as a **rapid evidence summary**, not a full systematic review, which may have introduced selection and appraisal bias. Only a subset of studies underwent double data-checking, and citation searches may have missed relevant publications.
All included studies were **retrospective, non-UK**, and based on **enriched datasets**, limiting generalisability. No **prospective or real-world trials** of AI in screening workflows were found.
Many studies lacked **predefined thresholds**, involved **manufacturer-linked data**, or were conducted in **laboratory settings**, reducing clinical applicability.
Overall, the evidence is **heterogeneous and methodologically weak**, and findings may **not directly translate to NHS breast screening practice.**

Conclusions
AI offers strong potential to enhance breast cancer screening through improved accuracy and workflow efficiency.
However, existing evidence—mainly **retrospective, non-UK, and based on enriched datasets—is insufficient for NHS adoption.**
Among CE-marked tools, **Transpara** shows greatest readiness, but **UK-based prospective validation** remains essential.
Future research should evaluate real-world performance, cost-effectiveness, and integration within NHS workflows to ensure AI delivers safe, equitable, and evidence-based benefits in population breast screening.

References:

[illegible]