NHS FELLOWSHIP IN CLINICAL AI · COHORT 4 · 2025–26

Enriching clinical registries with AI and Python

An on-premise NLP pipeline that reads renal documents, extracts structured phenotypes, and writes them back into Epic — deepening a glomerulonephritis (GN) registry built to power predictive risk models at the point of care.

Dr John Hartemink MBBS MSc BSc MRCP · NHS Fellow in Clinical AI · Epic Physician Builder

Manchester Institute of Nephrology & Transplantation, Manchester University NHS Foundation Trust · Division of Cardiovascular Sciences, University of Manchester

~4,000	5	52+	673
reports processed	document types	structured fields / report	GN registry cohort

00 The problem — and who it affects

A 180,000-patient registry, enriched by reading the prose that structured fields miss

180,000

kidney patients in one Trust's registry

Manchester University NHS Foundation Trust runs a live renal registry of **~180,000** patients. Care decisions across the whole population lean on it — so whatever the registry cannot see, the clinic cannot act on.

~10,000

legacy biopsy reports alone — every one prose

The registry is built from structured EHR fields, but the details that actually decide care — diagnosis, severity score, scarring, antibody staining — exist only as narrative text. Biopsy reports alone run to **~10,000 documents**, before letters, referrals and echoes. **Structured data alone misses them**, and re-typing by hand cannot scale.

NLP

+ Python — enrichment at population scale

The pipeline reads the unstructured reports and letters and pulls the exact datapoints the registry lacks, writing them back automatically. **~4,000 documents processed so far**, with thousands more to come — prose becoming structured data, enriching the registry across ~180,000 patients so every one can receive optimal, data-driven care.

01 From document to structured phenotype

Five source types → one Python pipeline → registry-ready data

DOCUMENTS IN — FIVE KINDS OF RENAL RECORD

Renal biopsy reports	Clinic letters	Referral letters	TTE / echo reports	Medication lists
Diagnosis, Oxford/MEST-C, ISN/RPS, IF, EM	Epic Reporting Workbench export · two-stage	Diagnoses, medications, summary	LVEF, valve function, chamber size	Drug, dose, form · KDIGO drug classes
LLM	LLM	HYBRID	REGEX	REGEX

STEP 1 Ingest — PDF & Excel; merged-cell recovery from Epic exports

STEP 2 Triage — Specimen class — renal, oncology, non-diagnostic; drops non-renal early

STEP 3 Split — Section detection — LM, IF, EM, diagnosis, impression

STEP 4 Extract — A local model reads the text. Any on-premise model drops into this slot — the pipeline does not change

STEP 5 Validate — Gold-standard scoring · clinician review · hallucination checks

STEP 6 Export — Structured JSON & CSV, ready for Epic

4,554 reports → 3,159 renal → 1,138 locally extractable · every model runs inside the Trust · no clinical text leaves MFT

02 Back into Epic — as registry metrics

Structured phenotypes become live population-health measures

Structured JSON (Pipeline output) → **External Data Manager** (Epic ingestion point) → **Registry Data Table** (Per-patient metric) → **Healthy Planet / Epic Radar** (GN registry dashboard)

A live platform, not a report

- **Continuously updated** — Every new biopsy report or clinic letter filed is read and lands in the registry. The cohort is current today — not an annual data return that is stale on arrival.
- **Clinician-ready inside Epic** — It opens in the record the team already works in. No separate login, no export request, no waiting on an analyst.
- **Used in routine care** — Clinic preparation, referral triage, therapy gaps and drug-safety monitoring — applied directly to the patient in front of you, not banked for later research.

Owned by the clinicians and the provider team. MFT builds it, holds it and governs it. It is not a vendor product and not a research silo — the people who use it are the people who decide what it does.

NLP-derived registry metrics

METRIC	SOURCE DOCUMENT
Oxford/MEST-C score	Renal biopsy report
ISN/RPS class (lupus nephritis)	Renal biopsy report
IF / EM findings	Renal biopsy report
Disease-activity status	Clinic letter
Treatment intent	Clinic letter

The registry it enriches

673 — GN registry cohort · 7 subtypes

267 — IgA nephropathy sub-cohort

0–223 — **GN Diagnosis Score** — multimorbidity encoding for risk stratification

GDMT — **Guideline-directed therapy** gap — KDIGO-indicated treatments not yet prescribed, surfaced as a dashboard alert

Worked example — IgA nephropathy · phenotype-guided therapy

Biopsy report — NLP extracts: IgA nephropathy · MEST-C M1 E0 S1 T0 C0 · mesangial IgA 3+, C3 3+

Registry — GN Diagnosis Score 37 · IgA sub-cohort (n=267)

GDMT foundation — RAASi + SGLT2i optimised · targeted-release budesonide

Suitability for novel agents — Strong mesangial C3 → complement-driven → candidate for a complement inhibitor (e.g. iptacopan) as the IgA pipeline arrives

03 Development pathway

From zero-shot prompting to fine-tuned, subspecialty-aware extraction

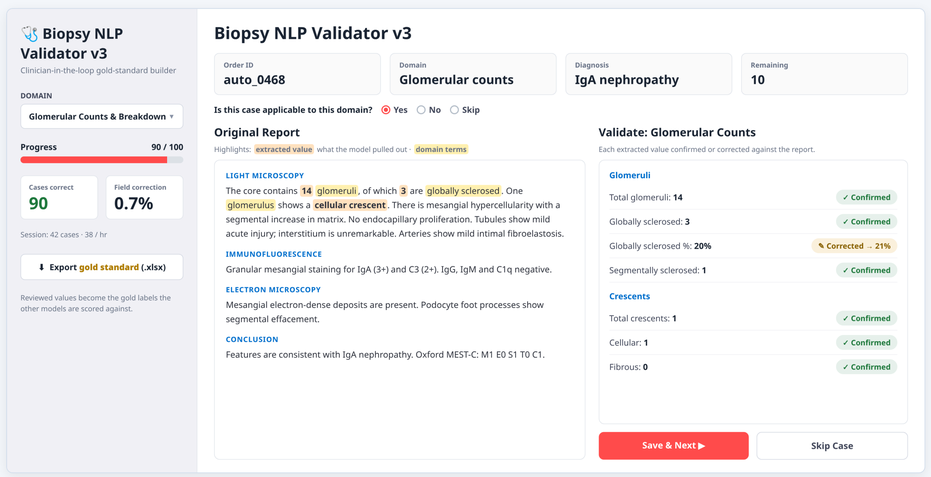
Zero-shot in GN — Open-weights · prompt-only · on-premise

Fine-tuned across natives — LoRA gold standard · all native disease

Subspecialty reporting — Transplant · Vasculitis · Paeds

04 Validated — the gold-standard engine

Clinician-in-the-loop review builds the gold standard the models are scored against



Biopsy NLP Validator v3. The reviewer confirms or corrects each field against the source; reviewed values become the gold standard the other models are scored against — field-level cite-or-retract.

Validation results — accuracy per model†

	Gemma-3	Phi-4	Qwen3-14B	Qwen3.6	MedGemma
Diagnosis primary dx	83.1%	88.3%	79.2%	93.5%	100%
Glomerular counts field-level	93.8%	93.2%	96.1%	96.2%	99.1%
IF intensity per-stain	85.5%	85.1%	81.5%	84.7%	88.3%
IFTA severity grade or %	66.7%	73.3%	80.0%	93.3%	93.3%
Vessels hyalinosis · intimal	88.2%	65.7%	86.3%	82.4%	99.0%

MedGemma built the reviewer-validated gold — the ceiling by construction. Of the swappable models, **Qwen3.6** — run on a MacBook via Mixture-of-Experts — leads on diagnosis and IFTA. All five are swappable on-premise models.

Clinical safety — on-premise inference · de-identified text only · no third-party model exposure · **DCB0129** hazard analysis & bow-tie safety case

What's next — scale up: a very large locally-hosted model (up to **405B** on 2× **NVIDIA DGX Spark**, 256 GB unified) · **scale out:** Mixture-of-Experts · 4-bit quantisation · LoRA — large models on the consumer hardware clinicians already own · multi-user validation · prospective evaluation to **CONSORT-AI / SPIRIT-AI**

05 From deep phenotype to prediction to the bedside

Planned extension — the enriched registry becomes the substrate for risk models

Deep phenotype — MEST-C · ISN/RPS · GN score · eGFR trajectory

Predictive model — Registry-based, trained externally

Risk score in the chart — Per-patient · updates with the registry

Personalised medicine — Therapy tailored to phenotype & risk

Target risk models: kidney-failure progression (KFRE-style) · GN flare & relapse · treatment response · trial-ready, phenotype-matched cohorts

06 The fellowship year — and what happens next

Built with the Trust's data team, the Epic analysts and the university

What was delivered

- **CKD registry live in Epic** — ~180,000 patients; GN dashboard specified across 18 clinical quality measures
- **NLP pipeline built and run** over 4,554 biopsy reports, extended to letters, referrals and echo
- **Five models benchmarked** against a clinician-corrected gold standard across five extraction fields
- **Clinical safety case** — CSO certification, DCB0129/DCB0160 hazard analysis and bow-tie
- **Mandate to lead** the new MFT AI Governance Panel

Who made it possible

A clinician cannot build this alone. The year worked because the people who own the data, the EPR and the method sat at the same table.

Prof. Sandip Mitra
Supervisor — Manchester Institute of Nephrology & Transplantation

Dr Anthony Wilson & the MFT Clinical Data Science Unit
Host AI team — deployment route and information governance

Ehsen Mumtaz
MFT Epic Analyst — registry and dashboard build

Dr Mohamed Elewa
Consultant Nephrologist — Manchester Institute of Nephrology & Transplantation

NHS Fellowship in Clinical AI, Cohort 4
Faculty and cohort — protected time, curriculum, structured supervision

Claude CLI (Anthropic)
AI pair-programming — pipeline build, tooling and analysis

Next steps after the fellowship

- **Fine-tune for immunofluorescence** — LoRA against the gold standard, aimed at the weakest field
- **GN dashboard go-live** with NLP-derived fields in front of clinicians
- **Algorithmic audit** — subgroup performance in the Databricks shadow validation
- **Prospective evaluation** to CONSORT-AI / SPIRIT-AI
- **Beyond one Trust** — expand to other Epic and non-Epic sites through the nephrology network; the pipeline is site-configurable
- **Deep learning** — dialysis and nephrology outcome-prediction projects built on the enriched registry
- **Computer vision** — renal histopathology using these extractions as ground truth; a pathology reporting assistant
- **Governance** — chair the new MFT panel that gates predictive-model go-live

On-premise inference · **De-identified** text only · **Zero** third-party model exposure · **DCB0129** clinical-safety framework · **Clinician-in-the-loop** validation

john.hartemink@mft.nhs.uk

† provisional — model figures from the current pipeline run; validation accuracy is single-rater (n=100)

Selected work: Hartemink J. *Digital Innovation in Nephrology: From Registries to AI* (PhD thesis, in preparation). · NHS Fellowship in Clinical AI, Cohort 4 (2025–26). · Supervisor: Prof. S. Mitra.