RadAugment: An AI-Enabled Competency Tracking and Analytics Platform for Radiology Trainees.

Dr Yogeshwar Gohil[1], Dr Rahul Rana[1]

2nd Annual Global AI Conference 2026
Human + machine: clinician-led AI for tomorrow's healthcare

Introduction:

- UK radiology trainee logbooks live in paper diaries, e-portfolio free-text, and personal Excel files. None of these identify patterns; they primarily store records.
- We propose that trainees keep meticulous logs but their value is not being maximally utilised.
- Even when logs exist, they are primarily descriptive of reporting numbers, not primarily of learning needs. The UK is short ~30% of the radiology workforce required for current demand; without high quality teaching, the knowledge of tomorrow's generation of radiologists is at risk.
- The current default is that consultants and trainees walk into reporting sessions without a shared/collaborative picture of what the trainee actually struggles with.

Within the confines of a shortage of consultants, we have to make every minute of consultant-trainee contact more targeted. We propose an AI-enabled architecture to solve this.

The Innovation - RadAugment Architecture:

Voice capture. Trainee dictates a 30-second note at the end of a reporting list.

↓

Clinical parsing. The LLM extracts (a) cases reported, (b) anatomical regions, (c) named clinical uncertainties (e.g. "Lisfranc", "occult scaphoid", "subtle pneumothorax").

↓

Socratic inquiry. For each uncertainty, the platform asks the trainee a targeted question, e.g. "What specific landmark on the AP foot would confirm Lisfranc alignment for you?" The trainee answers; learning needs are identified for review with a consultant.

↓

Quantitative extraction. Volume metrics (number of films), thematic tags, and trainee-authored action plans are stored as structured records.

↓

Longitudinal dashboard. Recurring themes (e.g. "scaphoid" flagged in 3 of 4 weeks) surface as bottlenecks at the top of the dashboard.

The Pilot:

- Participants: 5 UK radiology trainees, grades ST1-ST3, recruited across West Midlands.
- Scope: Plain film reporting only (deliberately narrow, as there is high case volume, well-defined anatomical regions, suitable for early validation).
- Duration: 4-week pilot.
- Logging protocol: Trainees dictated one voice note per reporting session (~30 seconds). The platform parsed each note and ran the Socratic loop thereafter
- Outcome measures:
  - Daily logging time (baseline self-reported vs. RadAugment-measured).
  - Uncertainties captured per trainee (count of distinct uncertainties surfaced).
  - LLM parsing accuracy (% of clinical terms correctly identified, judged against the trainee's intent).
  - Usability (post-pilot trainee score, 1-10).
- Baseline comparator: Each trainee self-reported their typical Excel/paper logging time before the pilot, then logged via RadAugment during the pilot.

Results:

331 Plain films logged.

↓ **55%** daily logging time (55s → 24s)

7.6/10 trainee usability score

44 insights uncovered

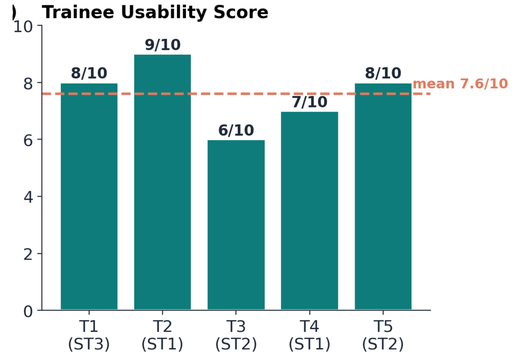

Figure one, a bar chart to demonstrate scores given by the piloting trainees to describe the usability of the platform. 0 = unusable, 10 = maximally UI-friendly

"When the tool asked me what my specific action plan was for the next equivocal scaphoid, it made me commit to a learning goal instead of just moving past the case." - ST1 Trainee

"Seeing that I flagged 'subtle alignment doubts' three times in two weeks was an eye-opener. Excel wouldn't show me that pattern; it would just show three 'foot radiograph' logs" - ST2 Trainee

Conclusions and next steps:

We propose 'RadAugment', an AI enabled strategy to maximise the yield of each trainee-supervisor exchange. Through AI enabled learning, the trainee still reflects, still authors their learning needs, and still takes ownership of the e-portfolio, but the platform performs the interview that a senior would otherwise have to perform unstructured.

Whilst the tool shows reductions in logging time and 44 'uncertainties' identified which would never have entered a logbook, the primary value of such a tool is making 'previously invisible' difficulties visible. Whilst any reporting domain with high case volume and recurring search strategies is a candidate for use with a similar tool, we recognise the limitations of this project:

- Small sample size (n = 5)
- Self-reported baseline; pre-pilot logging time was self-reported, not directly measured, providing room for error.
- Parsing accuracy 68%; domain-specific eponyms (Lisfranc, Salter-Harris, etc.) reduce accuracy.
- Scope deliberately narrow; Plain film only. Procedural skills excluded from this pilot.

Future directions:

- Further trials are required to validate this, on a larger scale, but provided this validation is achieved, we envision that such tools would be invaluable not just for all diagnostic radiology, but for procedural trainees who need to log learning needs for drains, biopsies, and US-guided procedures.

Acknowledgments: We would like to thank our pilot trainees. [1] Both authors contributed equally to the project. No conflict of interest to declare. This was an unfunded pilot.