

Matthew TH Lowry MD PhD^{1,2}, Jason Leake PhD¹, Janice Kiely PhD¹, Richard Luxton PhD¹, John McCormick MD¹
¹ South West Centre Digital Excellence, University of West of England, Bristol, UK;
² Department of Cardiology, Bristol Heart Institute, Bristol, UK

Ambient Voice Technology

- **Ambient Voice Technology (AVT)** combines speech recognition with generative AI to create structured medical documents
- The use of AVT has been shown to improve patient-clinician interaction, reduce appointment length and increase number of patients seen per shift in Accident and Emergency (A&E) (1)
- Regulatory requirements include the need for post market surveillance by the adopter (NHS) [DCB0160].

Challenges to Adoption

- Errors can occur in AVT summaries with the performance varying by model version. (2) This requires review of output by physicians and ongoing surveillance by healthcare institutions. Human review of outputs is the gold standard which carries a significant time and resource burden

Error	Rate (n = 12,999)	Severity >=major
Hallucinations	191 (1.47%)	84 (44%)
Omissions	1,712 (3.45%)	286 (16.7%)

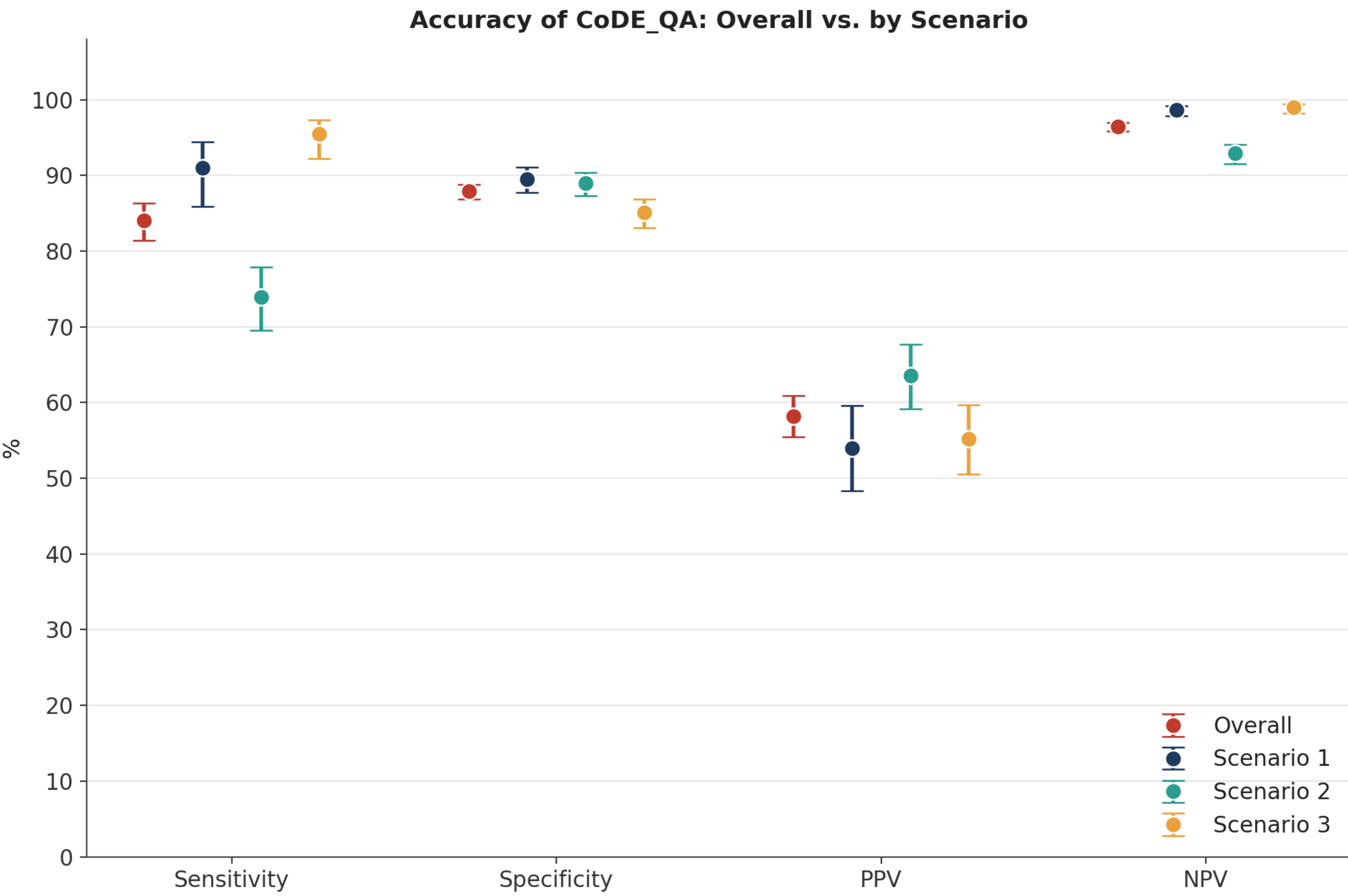
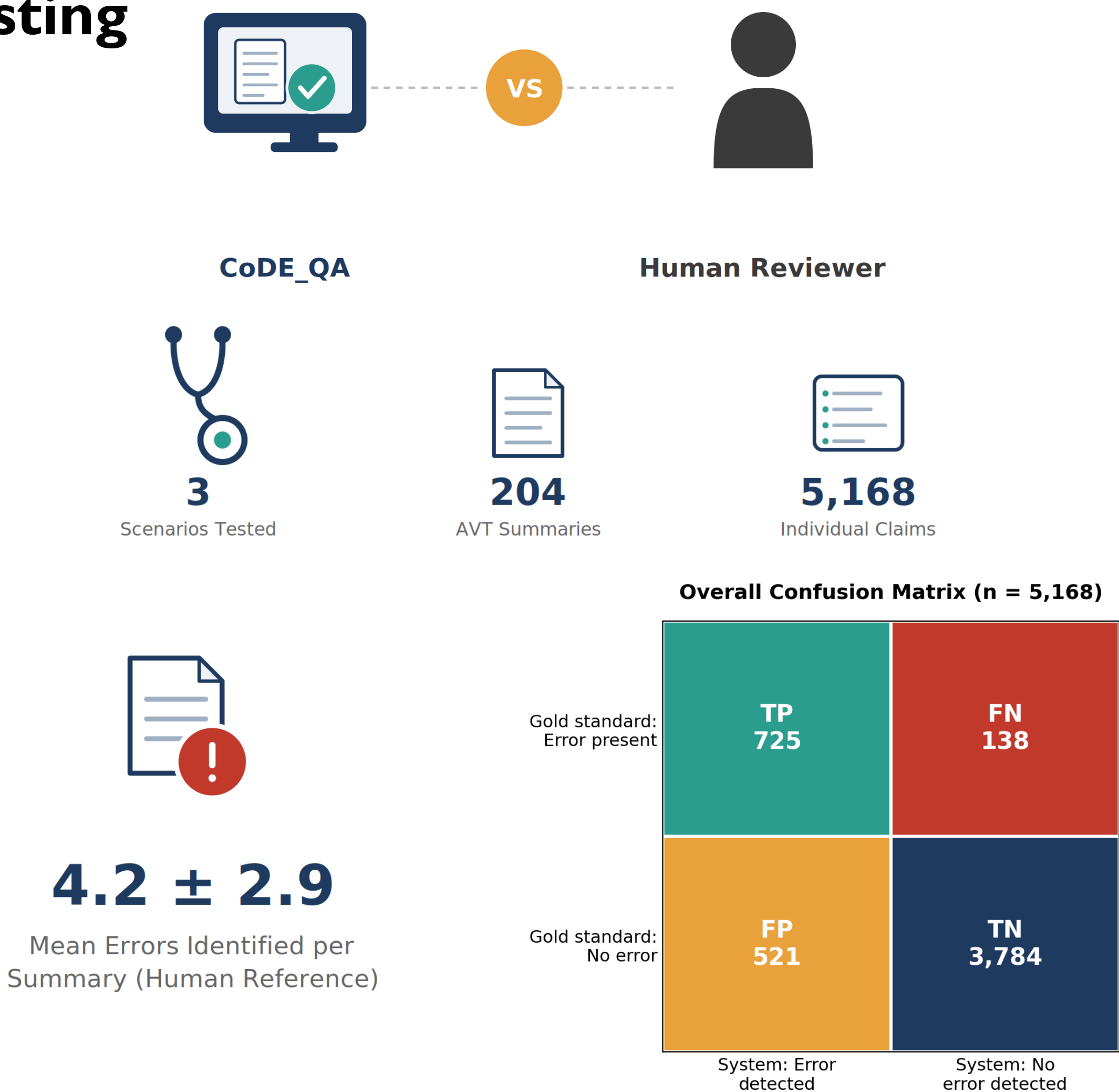
Aim: Design a provider agnostic system for automated performance monitoring of ambient voice technology

Design and Testing of CoDE_QA

Design

- CoDE_QA is a claims verification and semantic similarity comparison system which can identify errors in AVT produced clinical documents using the consultation audio transcript

Testing



Sens	Spec	NPV	PPV
84.0% (81.4-86.3%)	87.9% (86.9-88.8%)	96.5% (95.9-97.0%)	58.2% (55.4-60.9%)

Take Home Message

- AI scribes are being adopted faster than the evidence base or monitoring infrastructure can keep pace with.
- Hallucination and omission are inherent, measurable properties of LLM summarisation – safe adoption needs continuous surveillance
- CoDE_QA can identify errors in AVT summaries from the consultation audio transcript (omissions and hallucinations)
- CoDE_QA is desinged to support NHS England's Class I medical device and AVT Registry compliance requirements in practice.

Future Directions

- Expand test data set to increase scenario variety
- Assess error identification by clinical significance and likelihood of harm
- Fine-tune claim verification model to improve true positive and reduce false positive rate
- Transition our test system to a sandbox environment for longitudinal surveillance of live NHS AVT deployments.

References:
1. Great Ormond Street Hospital for Children NHS Foundation Trust. GOSH-led trial of AI-scribe technology shows 'transformative' benefits for patients and clinicians across London [Internet]. London: GOSH; 2025 Sep 4 [cited 2026 Jul 12]. Available from: <https://www.gosh.nhs.uk/news/researchgosh-led-trial-of-ai-scribe-technology-shows-transformative-benefits-for-patients-and-clinicians-across-london/>
2. Asgari, E., Montaña-Brown, N., Dubois, M. et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digit. Med.* 8, 274 (2025).