NHS England

Fellowship in Clinical AI

Dr Michael Mackay[1], Dr Ari Ercole[1,2]
1. Cambridge University Hospitals NHSFT, 2. University of Cambridge

NHS Cambridge University Hospitals NHS Foundation Trust

Deploying generative AI in Epic

Project

Epic's In-basket ART system uses OpenAI's GPT-4o large language model (LLM) to automatically draft text for a potential response to a patient's MyChart message based on the text of the message, trust-specific prompts, and information from the patient's record, such as current prescriptions and recent results[2]. Clinicians can review and revise the response before sending - or ignore it and draft their own reply from scratch.

Rationale

Generative AI can lift much of the paperwork burden that keeps clinicians glued to keyboards instead of patients. The reclaimed time means more face-to-face care, shorter workdays, and lower risk of burnout. The system produces standardized, legible text meaning consistency in tone and responses across users[2].

Review of experience of centre's who have deployed ART. Planned time-in-motion study pre- and post-deployment to quantify time saved and user satisfaction. Planned development of patient questionnaire for feedback about messaging

Developing SOP for monitoring post-deployment:

- Automated analysis of similarity between draft and sent messages using ROUGE, BERTscore, and Levenstein distance
- Monitoring for drift – changes in similarity over time suggesting changing context or practice
- Monitoring inter-user variability – monitoring for mis-use
- Regular audit of sample of messages by relevant specialists

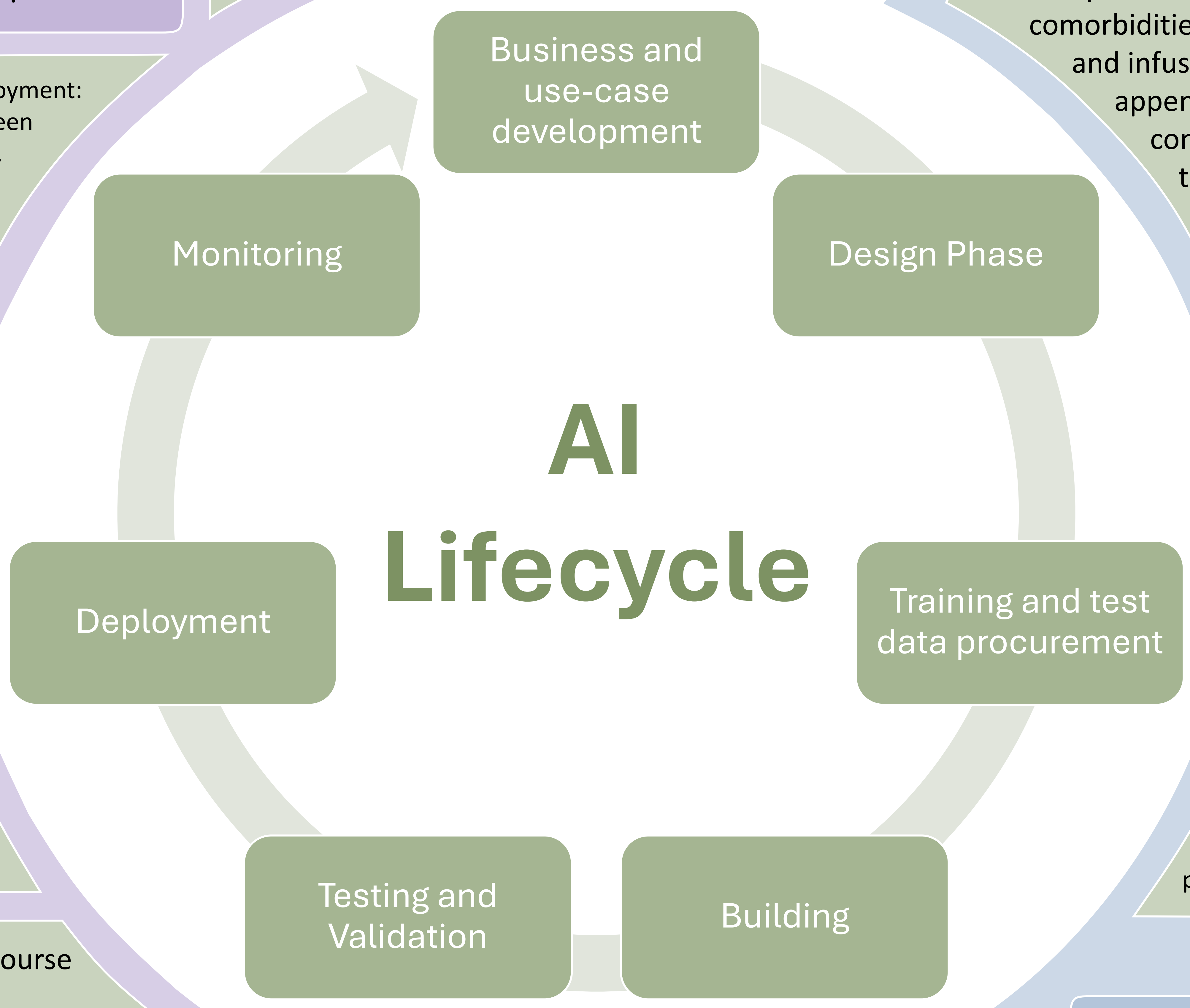
Completed data privacy impact assessment
Wrote clinical safety case and led hazard assessment workshop
Developing user SOP
Deployment planned as a phased rollout by department
Initial deployment planned in midwifery department

Completed Epic Physician Builder course
Developed pipeline for T&V:

- Initial iterative prompt engineering
- Specialty review with final prompt engineering
- Blinded evaluation of draft responses and real responses by target users
- Time in motion studies with human and automated evaluation of draft and sent responses

Next steps

- Blinded comparison study of draft vs midwife messages
- Limited Go-live within subset of midwifery department
- Time-in-motion studies pre and post-deployment
- Automated analysis of similarity metrics between draft and edited/sent messages
- Finalisation of monitoring SOP
- Roll-out to additional departments



Predicting ICP Augmentation in NCCU

Next steps

- Data cleaning and validation
- Data exploration and feature selection
- Replication of model with Epic dataset
- Testing and validation
- Integration into Epic docker container and deploy as custom model within Epic

The design of the model is informed by previous work on predicting TIL[1]. We utilise all data captured within the Epic EHRS, including comorbidities, vitals, lab results, medications and infusions. Variables are tokenised by appending their ordinal value, binned continuous value, or "not known" to the variable name, and then vectorised. For each day a set of vectors is then derived from the patient parameters and used to train a RNN to predict the TIL the following day.

CUH has overarching ethics approval allowing the use of data collected as part of routine clinical practice and held on their EHRS to be used anonymously for research. I created a project plan and completed the data request application which received committee approval, then worked with the analyst to identify the patient cohort in Epic and the relevant data to extract. 10% will be put aside as a held-out set.

Rationale

Another application of AI in healthcare outside clinical decision support is in operational support – the ability to look across multiple patients, wards or the whole hospital and predict/organise resource/capacity requirements. By predicting TIL ahead of time, we can inform advance contingency planning for capacity in overstretched ICU departments

Building is done within a secure azure stack within the CUH intranet ensuring no patient data leaves the trust. The model is built in python allowing integration into an Epic docker container to facilitate trialling the model within an Epic test environment as a custom AI model.

Project

When traumatic brain injury (TBI) patients are admitted to the intensive care unit (ICU), a core focus of their care is to protect and promote potential recovery in brain tissue by either preventing or mitigating raised intracranial pressure (ICP). The intensity of strategies to augment the ICP can be captured in a metric called the therapeutic intensity level (TIL). Previous work has developed a model utilising natural language processing (NLP) tokenisation and a recurrent neural network (RNN) to predict the TIL the following day[1]. This project aims to replicate this model utilising data from CUH's electronic health record system – Epic.

References

1. Bhattacharyay, S., van Leeuwen, F.D., Beqiri, E., Åkerlund, C.A., Wilson, L., Steyerberg, E.W., Nelson, D.W., Maas, A.I., Menon, D.K. and Ercole, A., 2025. TILTomorrow today: dynamic factors predicting changes in intracranial pressure treatment intensity after traumatic brain injury. *Scientific Reports*, *15*(1), p.95.
2. Gen AI Saves Nurses Time by Drafting Responses to Patient Messages. *https://www.epicshare.org/share-and-learn/mayo-ai-message-responses*

NHS Fellowship in Clinical AI
NHS Digital Academy Programme