Interaction of doctors with explainable RL decision support via behavioural readouts of eye-tracking

Myura Nagendran
Department of Surgery and Cancer
Imperial College London
St Mary's Campus, Praed Street, London, W2 1NY, UK
myura.nagendran@imperial.ac.uk

Paul Festor
Department of Computing
Imperial College London
180 Queen's Gate, South Kensington, London, SW7 2AZ, UK
paul.festor19@imperial.ac.uk

Matthieu Komorowski
Department of Surgery and Cancer
Imperial College London
St Mary's Campus, Praed Street, London, W2 1NY, UK
m.komorowski14@imperial.ac.uk

Anthony C. Gordon
Department of Surgery and Cancer
Imperial College London
St Mary's Campus, Praed Street, London, W2 1NY, UK
anthony.gordon@imperial.ac.uk

Aldo A. Faisal
Department of Bioengineering
Imperial College London
180 Queen's Gate, South Kensington, London, SW7 2AZ, UK
aldo.faisal@imperial.ac.uk

Abstract

Explainable reinforcement learning (XRL) is crucial for reinforcement learning (RL) algorithms within clinical decision support systems. However, most XRL evaluations have been conducted with non-expert users in toy settings. Despite the promise of RL in healthcare, deployment has been especially slow in part because of safety concerns which XRL might be able to attenuate. In our study, we observed doctors interacting with a clinical XRL in a high-fidelity simulated medication dosing scenario. Using eye-tracking technology, we analyzed these interactions across safe and unsafe XRL suggestions. We find that there cognitive attention devoted to XRL during unsafe scenarios is similar to during safe scenarios (despite doctors more frequently rejecting unsafe XRL suggestions). This suggests that XRL does not lie in the causal pathway for doctors to reject unsafe AI advice.

16th European Workshop on Reinforcement Learning (EWRL 2023).

1 Introduction

Healthcare is a high-stakes domain with recurrent decision-making in pursuit of a long term objective (typically maximising patient health/survival); in other words, the perfect setting in which to exploit the benefits of reinforcement learning (RL) algorithms. In practice, despite promising proof of concept papers [REF komorowski], there are no widely deployed clinical RL decision support systems even though these are most likely to be supportive rather than autonomous in the near future Festor et al. [2021].

Therefore, optimizing the interaction between healthcare practitioners and RL-driven AI clinical decision support systems (AI-CDSS) becomes vital for broad acceptance and influence, something that has been relatively overlooked to date van de Sande et al. [2021]. Explainable RL (XRL) has been proposed as a potential strategy by providing intelligible justifications for RL-driven recommendations to human users Barredo Arrieta et al. [2020]. Apart from fostering trust in RL, XRL has been proposed as a mechanism to prevent the inadvertent implementation of unusual or even detrimental AI advice Jia et al. [2022], Gordon et al. [2019], Antoniadi et al. [2021]. The urgency for this is heightened by the emergence of generative AI (such as large language models) which occasionally generate hallucinatory (and thus, if used clinically, unsafe) suggestions Lee et al. [2023]. Nonetheless, the extent to which XRL can serve as a guard against unwitting adherence to unsafe (i.e., hallucinatory) AI advice is still unclear Evans et al. [2022], Jacobs et al. [2021], Ghassemi et al. [2021].

When it comes to the real-world application of clinical XRL, there is a paucity of clinical evaluations involving XRL with specialist end-users, and even fewer in a high-fidelity setting Schoonderwoerd et al. [2021]. Current evidence indicates a weaker than expected correlation between physicians' actual prescribing behaviors and self-reported utility of XRL Nagendran et al. [2022]. Importantly, other investigators have highlighted that both self-reports and actual behaviors can only be recorded after the event Cao and Huang [2022], limiting the effectiveness of these retrospective metrics as part of a reinforcement learning feedback loop, in contrast to real-time clinical attention indicators like eye-tracking Ball and Richardson [2022], Harston and Faisal [2022]. This technique has been widely utilized in non-hospital scenarios to ascertain an individual's attention focus Auepanwiriyakul et al. [2018], Makrigiorgos et al. [2019], Ranti et al. [2020], Harston et al. [2021]. High-fidelity simulation environments offer an opportunity to investigate XRL in a setting that closely mirrors real clinical practice and is often used in medical education Cato and Murray [2010], Cook et al. [2011]. By incorporating eye-tracking into a high-fidelity setting, our methodology attempts to address the limitations of previous research (non-clinical participants, surrogate tasks, low fidelity environments) and gain a more accurate understanding of the clinician-XRL interaction dynamic.

In this work, we explored the influence of four distinct AI explanation types on clinicians within a high-fidelity simulation environment, as they performed a routine hospital task: determining the appropriate medication dosage for a patient after evaluation. Our aim was to quantify the impact of XRL on clinicians' prescription decisions, with a specific focus on whether doctors' attentional engagement (as measured by eye-tracking) differed between safe and unsafe AI scenarios.

2 Methods

Experimental Setup and XRL – Our investigation consisted of an observational analysis of how humans interact with AI within a simulated environment. Medical professionals were presented with one of six patient situations under two conditions: a recommendation from AI that was deemed safe or one that was potentially unsafe. The classification of safe and unsafe was based on exceptionally high or low prescriptions of fluids and vasopressors, as defined in prior research Festor et al. [2022]. The AI suggestions were artificially generated, with the main aim of our research being to assess the dynamics of interaction between healthcare professionals and AI. We constructed four unique explanations for the simulated AI system, all grounded in techniques we've utilized in reinforcement learning decision support systems. The first provided a natural language description of the Q-value difference between the recommended action and alternative actions. The second clarified the projected short-term changes in mortality following dosage adjustments as predicted by the AI. The third emphasized the five most influential aspects of the input data that steered the AI's recommendation. Finally, we identified the three most impactful training examples during the Q-learning process.

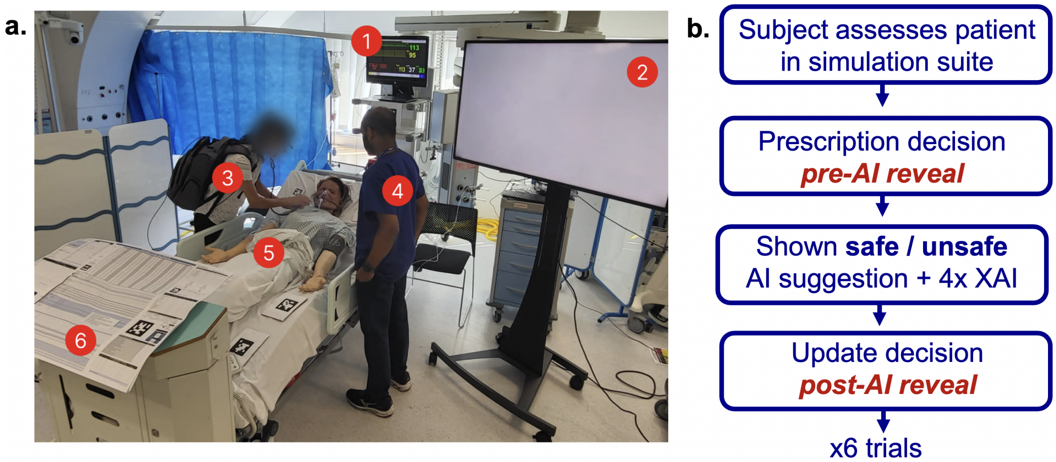

Figure 1: **Simulation suite (a) –** where numbers in the simulation suite refer to (1) vital signs monitor, (2) AI screen, (3) subject, (4) bedside nurse (played by experimenter), (5) high-fidelity patient mannequin, (6) bedside ICU data chart. Trial protocol (b)

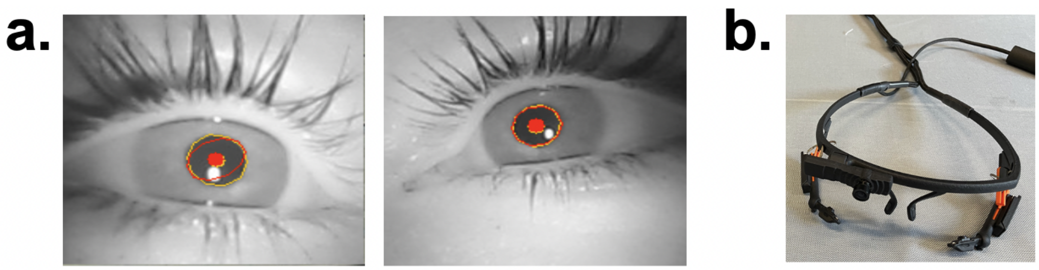

Figure 2: **Eye-tracking glasses (a) and pupil detection (b) –** Automatic pupil detection -> triangulates gaze position after calibrating software for each subject. Eye-tracking glasses have 3 cameras ('ego-centric' world-view camera plus one camera for each eye).

Eye-tracking for Gaze Detection – Eye-tracking was used to detect gaze, thereby determining clinicians' attention profile during simulations and their fluctuation. Participants wore unobtrusive, off-the-shelf eye-tracking glasses (Pupil Labs Core) featuring three cameras (Figure 2b), with the main camera capturing the wearer's viewpoint and the remaining two recording the eyes (Figure 2a). Pupil Labs software (Pupil Capture, version 3.5.7) used these cameras to identify the pupil and deduce the gaze direction (Figure 3a). Prior to the experiment, a two-stage 2D calibration exercise was conducted. Eye-tracking glasses were connected to a laptop (Lenovo Thinkpad) worn in a lightweight backpack, allowing unrestricted movement.

We delineated four primary regions of interest (ROIs) (Figure 1a): the patient mannequin, the vital signs monitor, the ICU data chart, and the AI display screen, the latter having four sub-regions tied to the XRL types. Using pre-set QR codes (April tags, Figure 3a), ROIs were defined during post-processing. This allowed analysis of gaze-time per ROI, fixation count per ROI, and blink rate per minute per ROI - an indicator of concentration level.

We further devised a distinctive method for gauging behavioural attention that adjusts for the percentage of the visual field an ROI occupies. To illustrate, if an ROI constitutes 50% of the visual field for half the time, based on randomness alone, the gaze should land within the ROI 25% of the time. By comparing this 'random gaze' figure of 25% with the actual gaze proportion, we can calculate a surrogate measure of the relative importance of ROIs, contrasting the rates of random and actual gaze into a ratio. The higher the ratio, the more purposeful the attention is on any given ROI (versus chance gaze).

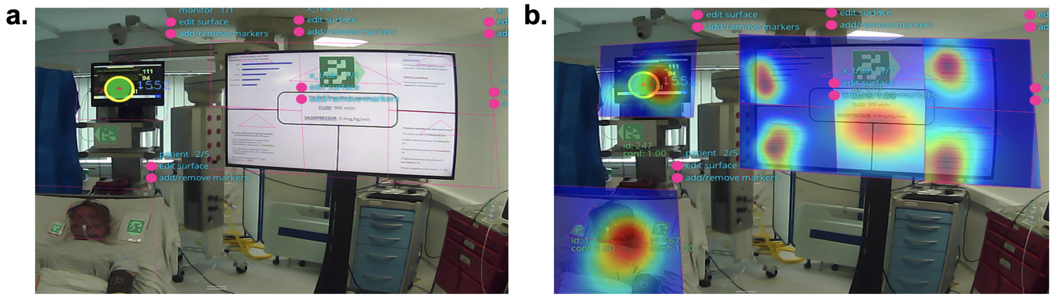

Figure 3: **Post-processing of eye-tracking data -** Left-hand image shows bounding boxes around regions (surfaces) of interest (ROIs). Right-hand side shows gaze density per ROI as heatmaps.

Simulation Experiment – Participants first underwent a standard experiment briefing and completed a pre-experiment questionnaire regarding AI perceptions and demographic information. After familiarising themselves with the simulation suite and performing eye-tracking calibrations, they began the simulated scenarios. An individual role-playing an ICU bedside nurse assisted. The doctors evaluated six simulated ICU patients with sepsis. Each of the six scenarios required clinicians to perform an assessment, including data review and patient examination. Subsequently, they were asked by the nurse to provide their prescription for fluid and vasopressors for the next hour, indicate their confidence in the prescription, and state if they would seek higher advice or a second opinion. Afterward, AI recommendations and explanations were displayed near the patient bedside, prompting doctors to affirm or amend their prescriptions and reassess their answers regarding confidence and the need for senior advice (Figure 1b).

Subject recruitment – ICU doctors were recruited via targeted promotion and convenience sampling within a local hospital area. Eligibility requirements included: (i) currently practising as a doctor, (ii) possessing at least two months of adult ICU experience, (iii) current ICU involvement or ICU employment in the previous six months. Participants received compensation, and each session lasted around 60 minutes. The local research governance team at our institution and the UK Health Research Authority approved the study.

The previous sentence has been anonymised to allow for a blind peer-review process. If accepted, information about the institutions and approval reference number will be filled in on the camera-ready version.

3 Results

Cohort recruited – So far, 16 ICU doctors with eye-tracking data available were included (11 male, 5 female). Mean doctor age was 33 years (standard deviation (SD) 6 years). Mean ICU experience was 3.7 years (SD 4.1 years, range 2 months to 14 years).

Eye-tracking metrics on ROIs – There were more gaze fixations for the AI suggestion during unsafe scenarios but this was only significant for ICU doctors (>1 year experience in ICU, p=0.007, Figure 4). There were no significant differences in number of gaze fixations between the different XRL modalities for either safe or unsafe scenario, regardless of doctor seniority.

Mean blink rate was lowest for the ICU chart (6.0 blinks per minute (bpm), SD 4.5), similar for both patient and vital signs monitor (mean 14.4 bpm and 14.4 bpm, SD 9.5 and 9.1 respectively) and notably higher for the AI screen (mean 19.3 bpm, SD 11.4). When comparing all conventional clinical ROIs (chart, patient, monitor; blue bars in Figure 5) to all AI ROIs (including XRLs; orange bars in Figure 3), there was a significantly lower mean blink rate on the conventional clinical ROIs than the AI ROIs (11.6 bpm vs. 24.0 bpm, p=0.005).

For every ROI except the patient mannequin, there was a significantly higher actual gaze proportion than random chance gaze proportion (p<0.001 for all seven comparisons). For the major ROIs (AI screen, ICU chart, vital signs monitor, patient) the ratio of actual gaze to random gaze was 6.2, 1.6, 12.9 and 1.3 respectively. For the XRL ROIs (training examples, Q-value difference, mortality, feature

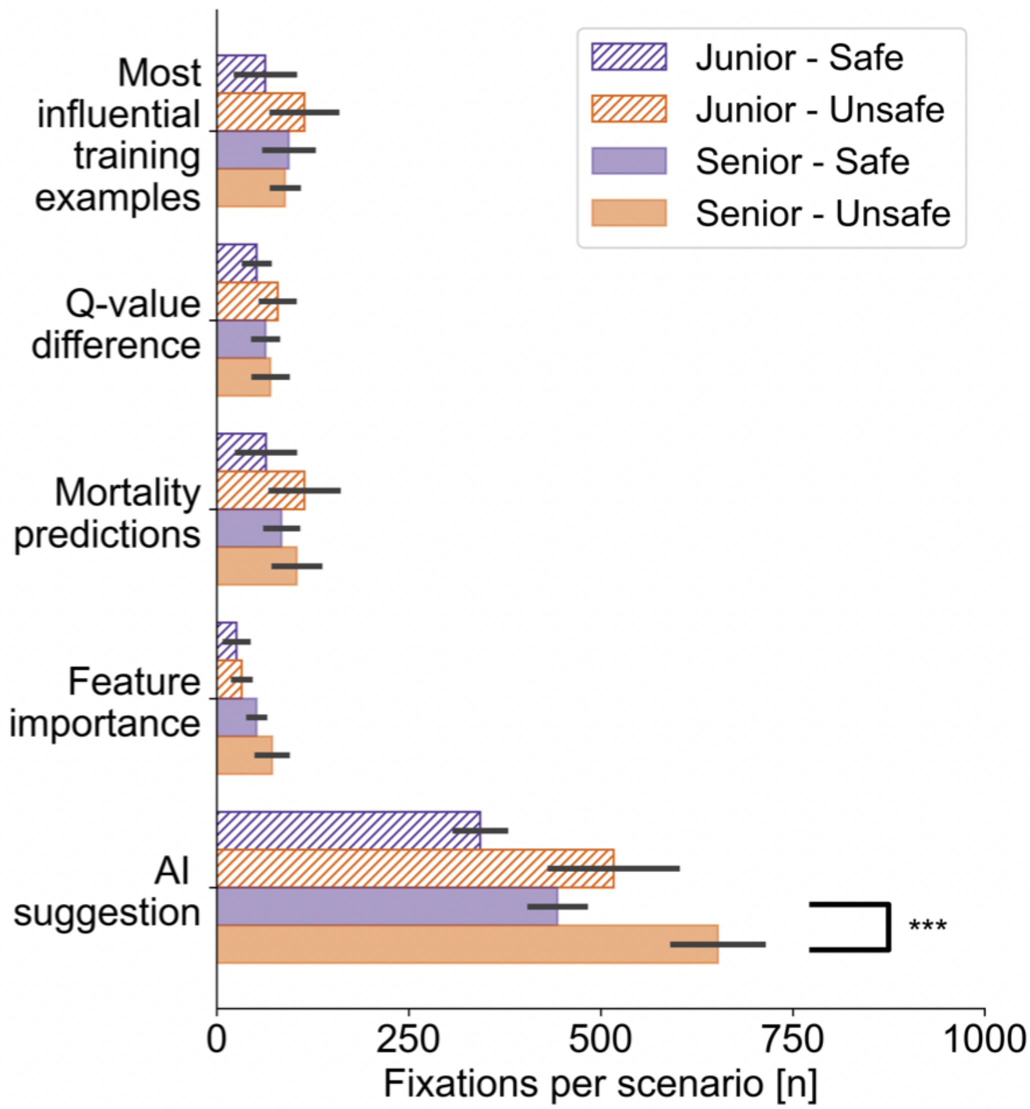

Figure 4: **Fixations per scenario by safety status of scenario and experience level of doctor –** Mean and SEM error bars.

importance) the ratio of actual gaze to random gaze was 5.4, 3.9, 4.9 and 3.1 respectively (see Figure 6).

Self-reported XRL usefulness – Self-reported data on the utility of XRL was only available for 10 of the 16 subjects (Figure 7). The overall mean post-experiment usefulness rating for the XRL was 3.0 (SD 1.1) on a zero to four scale with higher value implying the XRL was more useful. The training examples explanation was the only one of the four to be rated significantly lower than the overall rating for explanations in general (mean 1.0, SD 1.1, p<0.001). When comparing the 'objective' marker of how many fixations there were on the four different types of XRL to the 'subjective' marker of how clinicians rated the usefulness of the four XRLs, we found no significant correlation for any of the four XRLs.

Adherence to AI suggestions among doctors – We defined adherence to AI as the distance between a doctor's final prescription (having had the opportunity to view the AI suggestion) and the value of the AI suggestion for any given trial/scenario (higher distance suggesting that the doctor was less

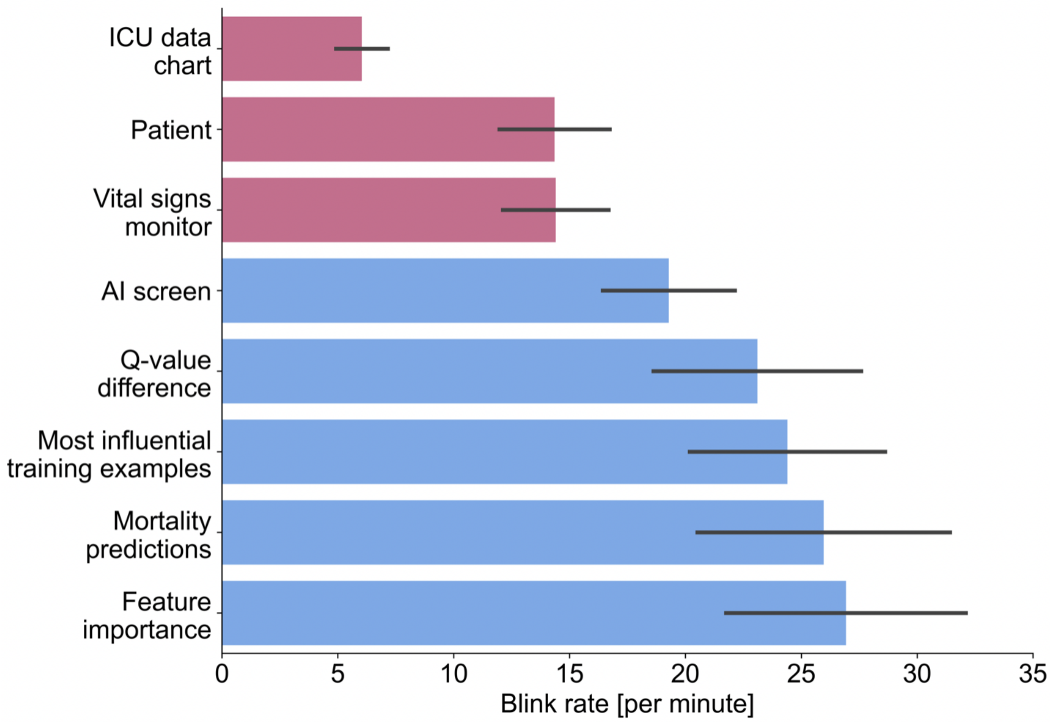

Figure 5: **Blink rate by region of interest –** Mean and SEM error bars. Blue bars are traditional clinical surfaces while orange bars are AI / XAI surfaces.

adherent to AI and vice versa). There was no evidence of correlation between eye-tracking metrics (blink rate or number of gaze fixations) and AI adherence regardless of safety status or drug. Nor was there a significant association between number of fixations specifically on XRL ROIs and drug (either fluid or vasopressor) for either AI scenario (safe or unsafe).

4 Discussion

Our study contributes several insights to our understanding of clinician interaction with XRL decision support tools. We demonstrated the efficacy of gaze fixations and blink rate as proxy attention indicators in a high-fidelity simulation, with the real-world clinical application pending less intrusive eye-tracking hardware and privacy considerations. We found no significant attention increase towards any explanation type when dealing with unsafe versus safe AI suggestions, challenging the assumption of heightened reliance on explanations in unsafe scenarios. There was no correlation between self-rated explanation usefulness and attention received, indicating that self-reports alone may not sufficiently evaluate XRL tools. Blink rate suggested less cognitive effort was required to interpret the AI data compared to the ICU chart. Lastly, no consistent links were found between eye-tracking metrics and variations in clinical practice or adherence to AI suggestions.

These findings must be viewed considering several limitations. The simulation suite couldn't fully emulate a real hospital environment's complexity, such as dynamic patient examination or team interactions. Real-world experiments require substantial sample sizes to standardise, making simulation studies crucial in exploring human-AI dynamics. Our small sample size could limit the validity of certain comparisons. Also, the categorisation of AI suggestions into safe or unsafe creates an arbitrary boundary on a continuous spectrum. Variations in explanation format could confound comparisons: some were primarily graphical, and others were text-heavy, potentially confounding comparisons between explanations.

Despite these limitations, our findings offer key insights for optimizing XRL-based medical decision support tools. We examined the assumption that explanations should help users reject poor AI advice,

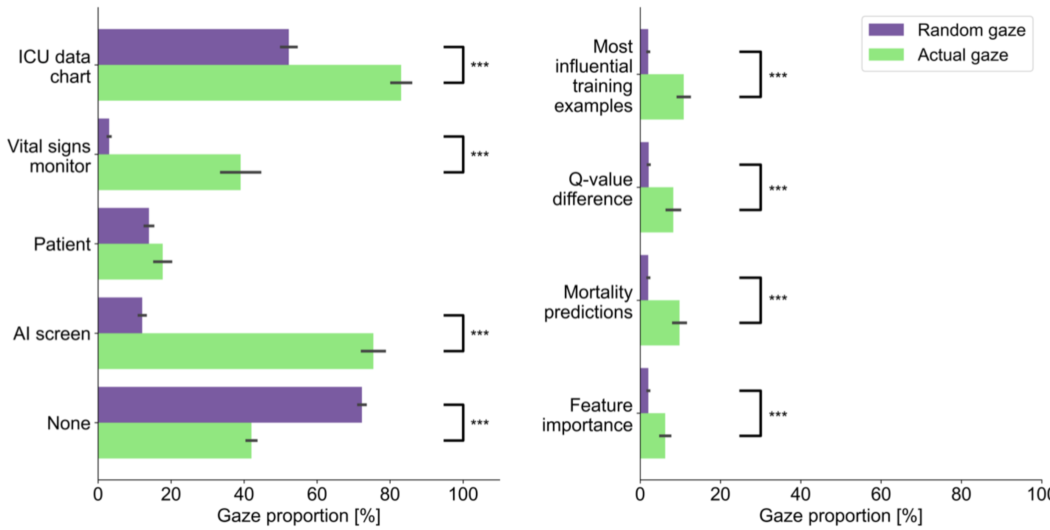

Figure 6: **Gaze proportion per ROI –** Mean and SEM error bars. Random gaze is the proportion expected based on the area occupied by an ROI within the visual field (i.e. if ROI takes up 50% of the screen, 50% of the time, we would expect gaze to randomly fall within it 25% of the time). Actual gaze is the proportion observed during the experiment.

and corroborated that an increased rejection rate of unsafe advice wasn't driven by higher reliance on, or attention to, explanations. Evidence from other studies suggests a breakdown in the process of discarding unsafe advice, indicating potential automation bias Shafti et al. [2022]. The risk of such automation bias is also well documented in other medical investigations Micocci et al. [2021], Panigutti et al. [2022]. Another piece of evidence is an experiment assessing a mental health drug decision support tool, where explanations failed to prevent clinical users from following intentionally subpar AI recommendations Jacobs et al. [2021]. Our study corroborates these findings.

The use of eye-tracking in AI-user studies remains limited. A notable example by Cao and colleagues found a positive association between gaze percentage on the AI suggestion and perceived user reliance and agreement with AI suggestions, but not with perceived trust Cao and Huang [2022]. Similarly, we found no correlation between subjective explanation ratings and gaze fixations on the AI explanation. While eye-tracking may form the basis of a real-time feedback loop for human-AI interactions Cao and Huang [2022]., our results caution that we must first establish reliable eye movement patterns to accurately categorize users and predict their AI interactions.

5 Conclusion

In summary, our results indicate that eye-tracking is a viable technique to assess clinicians' engagement with reinforcement learning explanations (XRL). We illustrate that clinicians' reactions to safe and unsafe AI recommendations are distinctly different. Yet, the absence of a 'rescue' effect presented by XRL is of importance to note when designing clinical XRL systems. Our insights emphasize the necessity for future AI decision-support tools to customise not just their recommendations, but also their interaction style and explanation delivery for clinician users.

Acknowledgments If needed, acknowledgments can be included right before the references in the camera-ready version. They must be hidden in the submitted paper.

Declarations An earlier version of this work is also under review at the 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH), part of the International Conference on Machine Learning (ICML) 2023.

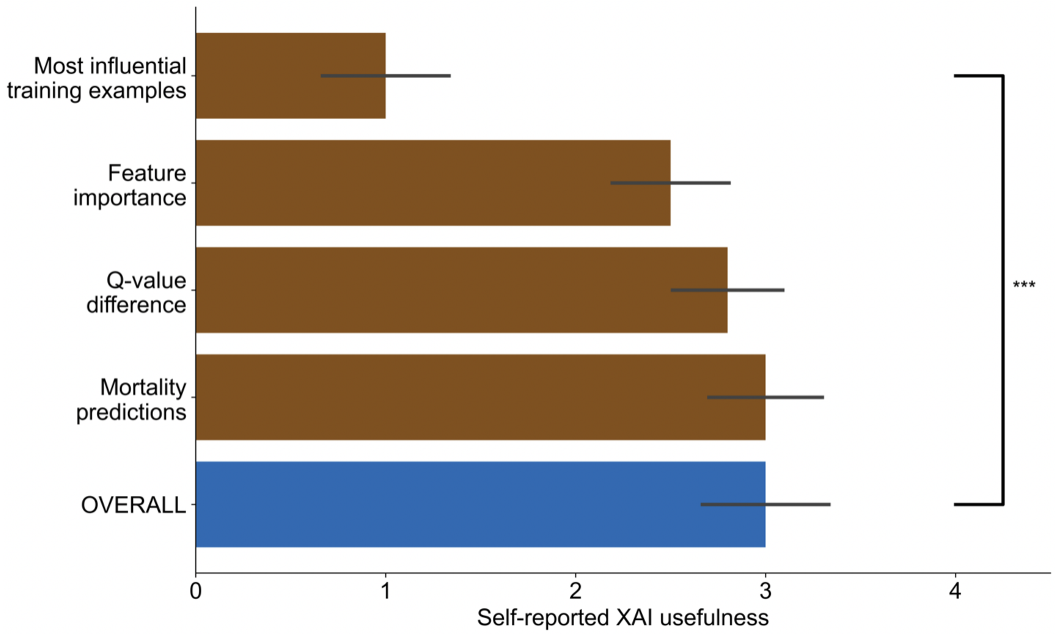

Figure 7: **Usefulness rating for each type of explanation as well as overall rating –** Mean and SEM error bars. The only significantly different explanation type compared to overall was the 'most influential training examples'.

References

AM Antoniadi, Y Du, Y Guendouz, L Wei, C Mazo, BA Becker, et al. Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: A systematic review. *Appl Sci*, 11(11), 2021.

C Auepanwiriyakul, A Harston, P Orlov, A Shafti, and AA Faisal. Semantic fovea: real-time annotation of ego-centric videos with gaze context. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research Applications*, page Article 87, Warsaw, Poland, 2018. Association for Computing Machinery. URL `https://doi.org/10.1145/3204493.3208349`.

LJ Ball and BH Richardson. *Eye Movement in User Experience and Human–Computer Interaction Research*, pages 165–83. Springer US, New York, NY, 2022. URL `https://doi.org/10.1007/978-1-0716-2391-6_10`.

A Barredo Arrieta, N Díaz-Rodríguez, J Del Ser, A Bennetot, S Tabik, A Barbado, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf Fusion*, 58: 82–115, June 2020.

S Cao and CM Huang. Understanding user reliance on ai in assisted decision-making. In *Proc. ACM Hum.-Comput. Interact*, volume 6, page Article 471. Association for Computing Machinery, 2022.

DL Cato and M Murray. Use of simulation training in the intensive care unit. *Crit Care Nurs Q*, 33(1):44–51, March 2010.

DA Cook, R Hatala, R Brydges, B Zendejas, JH Szostek, AT Wang, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*, 306(9):978–88, September 2011.

T Evans, CO Retzlaff, C Geißler, M Kargl, M Plass, H Müller, et al. The explainability paradox: Challenges for xai in digital pathology. *Future Gener Comput Syst*, 133:281–96, August 2022.

P Festor, I Habli, Y Jia, A Gordon, AA Faisal, and M Komorowski. Levels of autonomy and safety assurance for ai-based clinical decision systems. In I Habli, M Sujan, S Gerasimou, E Schoitsch, and F Bitsch, editors, *Computer Safety, Reliability, and Security SAFECOMP 2021 Workshops*, pages 291–6, Cham, 2021. Springer International Publishing.

P Festor, Y Jia, AC Gordon, AA Faisal, I Habli, and M Komorowski. Assuring the safety of ai-based clinical decision support systems: a case study of the ai clinician for sepsis treatment. *BMJ Health Care Inform*, 29 (1), July 2022.

M Ghassemi, L Oakden-Rayner, and AL Beam. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*, 3(11):e745–50, November 2021.

L Gordon, T Grantcharov, and F Rudzicz. Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg*, 154(11):1064–5, November 2019.

JA Harston and AA Faisal. *Methods and Models of Eye-Tracking in Natural Environments*, pages 49–68. Springer US, New York, NY, 2022. URL `https://doi.org/10.1007/978-1-0716-2391-6_4`.

JA Harston, C Auepanwiriyakul, and A Faisal. Prediction of visual attention in embodied real-world tasks. *J Vis*, 21(9), September 2021.

M Jacobs, MF Pradier, THJ McCoy, RH Perlis, F Doshi-Velez, and KZ Gajos. How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Transl Psychiatry*, 11(1), February 2021.

Y Jia, J McDermid, T Lawton, and I Habli. The role of explainability in assuring safety of machine learning in healthcare. *IEEE Trans Emerg Top Comput*, 10(4):1746–60, December 2022.

Peter Lee, Sebastien Bubeck, and Joseph Petro. Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine. *New England Journal of Medicine*, 388(13):1233–1239, 2023. doi: 10.1056/NEJMsr2214184.

A Makrigiorgos, A Shafti, A Harston, J Gerard, and AA Faisal. Human visual attention prediction boosts learning performance of autonomous driving agents. *ArXiv Prepr ArXiv190905003*, 2019.

M Micocci, S Borsci, V Thakerar, S Walne, Y Manshadi, F Edridge, et al. Attitudes towards trusting artificial intelligence insights and factors to prevent the passive adherence of gps: A pilot study. *J Clin Med*, 10(14), July 2021.

M Nagendran, A Gordon, and A Faisal. Impact of xai dose suggestions on the prescriptions of icu doctors, 2022. URL `https://2022.ccneuro.org/proceedings/0000328.pdf`.

C Panigutti, A Beretta, F Giannotti, and D Pedreschi. Understanding the impact of explanations on advice-taking: a user study for ai-based clinical decision support systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, page Article 568, New Orleans, LA, USA, 2022. Association for Computing Machinery. URL `https://doi.org/10.1145/3491102.3502104`.

C Ranti, W Jones, A Klin, and S Shultz. Blink rate patterns provide a reliable measure of individual engagement with scene content. *Sci Rep*, 10(1), May 2020.

TAJ Schoonderwoerd, W Jorritsma, MA Neerincx, and K van den Bosch. Human-centered xai: Developing design patterns for explanations of clinical decision support systems. *Int J Hum-Comput Stud*, 154:102684, October 2021.

A Shafti, V Derks, H Kay, and AA Faisal. The response shift paradigm to quantify human trust in ai recommendations. *ArXiv Prepr ArXiv220208979*, 2022.

D van de Sande, ME van Genderen, J Huiskens, D Gommers, and J van Bommel. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med*, 47(7): 750–60, July 2021.

A Appendix - Standardised experiment briefing to participants

This simulation experiment aims at studying how clinicians might interact with an AI decision support tool for patients with sepsis. The AI takes patient variables as input and outputs a treatment recommendation for both fluid and vasopressor. We have some evidence of effectiveness of the AI you will interact with on retrospective data from an American dataset and further validation on retrospective data from the Netherlands. However, there has not yet been prospective evidence of either effectiveness or safety.

You will conduct a brief ward round review of 6 ICU patients with sepsis. Between each patient, you will exit the room and be called in again to see the next patient by the nurse.

For each patient, you will first be asked:

- Treatment prescriptions for:
- Fluid in ml for the next hour
- Vasopressor in mcg/kg/min for the next hour
- Your confidence in the prescription on a scale from 1 (low) to 10 (high)
- Whether or not you want to get advice from another doctor / senior doctor

You will then be shown the AI treatment recommendation on a digital screen. The screen contains the AI dose suggestions in the middle (for fluid and vasopressor) along with 4 explanations for the suggestions (one in each corner)

You will then be asked:

- To what extent you agree with the AI suggestion on a scale from 1 (strongly disagree) to 5 (strongly agree)
- Whether you wish to adjust your treatment prescription for:
  - Fluid in ml for the next hour
  - Vasopressor in mcg/kg/min for the next hour
- Whether your confidence in your prescription (on a scale from 1 [low] to 10 [high]) has changed as a result of seeing the AI suggestion
- Whether or not you want to get advice from another doctor / senior doctor
- If the AI suggested treatment was to be administered to the patient, would you act to stop the administration?

We will now show you an example of the AI screen that you will encounter in the experiment. You can see that there is an AI suggestion in the centre of the screen which is how much fluid and vasopressor the AI recommends over the next hour.

Around the corners of the screen there are four different types of explanation which can be thought of as the AI trying to convey the rationale for its suggested doses. The mortality change explanation conveys information on what the AI predicts will happen to the overall mortality risk in the short-term based on potential dose increases or decreases. The most influential training examples explanation conveys which three training cases were most helpful for learning the current suggestion in the same way that we might base our own treatment choices on previous notable cases we learnt from. The feature importance explanation conveys which were the top five features (or items in the data) that were most useful to the AI in generating its current suggestion. The 'AI treatment options gap' explanation conveys how much one treatment strategy looks superior compared to alternatives. If all potential options are similar (i.e. a low gap) then it suggests that the AI has near equipoise for options and you may wish to use your own judgement more strongly (as you will have additional information from examining the patient for example). However, if the gap is high it suggests that the AI has identified one particular treatment strategy as superior to the alternatives and therefore it would be worth considering this recommendation more strongly than with a low gap recommendation.

B Appendix - AI explanations used

Q-value difference – This approach leverages the fact that once training is complete, the reinforcement learning (RL) Q table will contain Q values for any given state-action pair. The optimal action is the one with the highest Q value for any given state. However, the difference between this highest Q value and the alternatives might be very small or large. If large, then there is much higher anticipated value from following the recommended action compared to an alternative. This is dichotomised arbitrarily from a continuous value to make interpretation more simple for non-AI users.

Mortality predictions – This approach leverages the fact that mortality can be predicted for any given state in the RL state space. Therefore the impact of different dosing strategies that might result in transition to alternative states with different predicted mortalities can be displayed to the subject to highlight how alternate strategies might change the risk of death.

Feature importance – This approach leverages the fact that the state space for RL based sepsis algorithms is commonly constructed using a k-means clustering algorithm to enable dimensionality reduction. After the algorithm converges, the cluster centroids represent the average feature values for patients in a particular state/cluster. A new patient would be assigned to the state/cluster that minimised the distance from their feature values to the respective cluster centroid. Intuitively, with often over 40 features, some features will be closer to the cluster centroid value than others for any patient assigned to a given state. This is exploited to rank features in terms of their proximity to the cluster centroid (or average state feature values) given that the archetypal patient for whom an RL agent policy action most applies is a patient who is most typical of that state. So subjects can be shown the top five ranked features contributing to state assignment.

Most influential training examples – This approach leverages the fact that the difference in Q values for any given state-action pair between one iteration of Q-learning and the previous iteration reflects how valuable the currently seen training episode is for learning the optimal action for any given state. This is similar to an instance-based explanation used in deep learning imaging XAI where the explanation consists of showing similar image instances from the training instance to explain why a particular image classification has been made.

- Create empty Q and 'Q-difference' tables (both indexed by $Q(S, A)$ tuple).
- For 500,000 episodes:
  - Select a random training episode:
    - For each time-step:
      - Perform Q-learning.
      - Check the Q differences table $\rightarrow$ Is the Q difference (i.e., the difference between old and updated new Q values) from this step among the top 3 for this state-action tuple?
      - If so $\rightarrow$ Update the differences table with the ID for this episode.
- End with a dictionary of the top 3 influential episodes per $Q(S, A)$ tuple.