

# Intervention-Aware Clinical World Model for Post-Op Outcome Forecasting in Cardiology

Yunsung Chung<sup>1</sup>, Yingshuo Liu<sup>1</sup>, Abboud F. Hassan<sup>1</sup>, Han Feng<sup>1</sup>, Mary M. Maleckar<sup>1,2</sup>, Nassir Marrouche<sup>1</sup>, and Jihun Hamm<sup>1\*</sup>

<sup>1</sup> Tulane University, New Orleans, LA 70118, USA  
{ychung3,jhamm3}@tulane.edu

<sup>2</sup> Simula Research Laboratory, Oslo, Norway

**Abstract.** Many clinical prediction models treat post-intervention outcomes as a one-step mapping from baseline measurements to a future endpoint. However, recovery after a procedure often unfolds as an irregular trajectory: clinical observations, medication changes, repeat interventions, and physiological measurements are recorded asynchronously and can change risk assessment over time. We propose an *intervention-aware clinical world model* that represents each patient with a structured latent state and evolves it through time-ordered post-intervention events. The model first encodes baseline imaging into a 3D spatial latent state. It then updates this state using procedural context, static covariates, elapsed time, and peri-event physiological embeddings. Follow-up imaging provides training-only supervision through a latent forecasting objective. We apply the framework to atrial fibrillation ablation. During the 90-day recovery window, irregular post-procedure records provide clinically meaningful evidence for long-term recurrence risk. In repeated internal cross-validation on DECAAF-II, our model achieves AUROC 0.756 and AUPRC 0.777 for recurrence prediction. It also achieves a scar-extent MAE of 2.971 percentage points without requiring follow-up MRI intensities at inference. The learned state supports recurrence-risk queries at different horizons and retrospective input editing of blanking-period records.

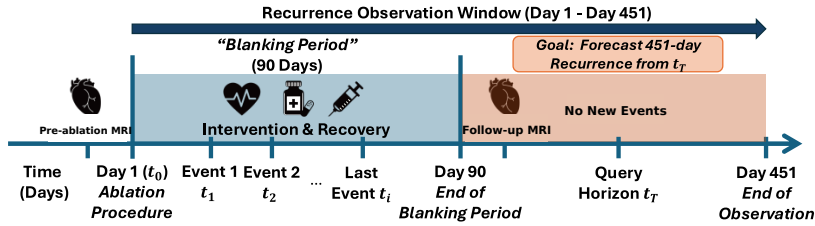
**Keywords:** Medical World Models · Clinical State-Space Modeling · Post-Intervention Forecasting · Multimodal Learning · Anytime Risk Updating.

## 1 Introduction

Many medical imaging models map a baseline scan directly to a future endpoint [7]. Yet recovery is dynamic: medication changes and follow-up interventions occur at irregular times and can change risk estimates. Models should therefore update predictions as longitudinal evidence accrues.

---

\* Corresponding author.



**Fig. 1. Clinical timeline:** Irregular interventions during the 90-day blanking period necessitate dynamic risk updating rather than a single-shot baseline prediction.

We study atrial fibrillation (AF) ablation. Pre-ablation LGE-MRI captures atrial fibrosis and anatomy. During the 90-day clinical “blanking period,” medication changes, electrical cardioversion, and occasional repeat procedures are recorded at irregular times [5]. These events are rarely modeled with baseline imaging. We forecast recurrence within 451 days using only records available by a selected query horizon (Fig. 1).

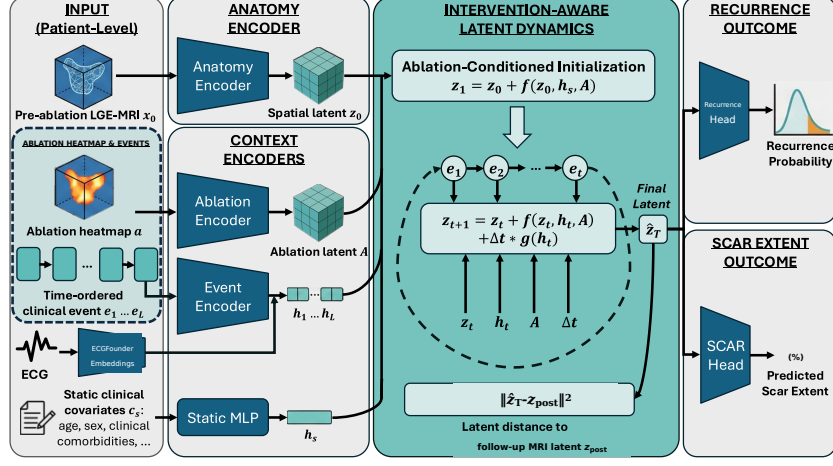
To address this gap, we introduce an *intervention-aware clinical world model*. It represents atrial anatomy in a spatial latent state, then updates the state with procedural geometry, static covariates, elapsed time, and pre-event ECG embeddings as blanking-period events occur. Rather than returning one fixed score, it provides retrospective risk estimates at selected horizons.

**Contributions.** We make three contributions:

1. An intervention-aware latent clinical world model that evolves a 3D anatomical state conditioned on ablation geometry, irregular post-procedural events, and ECG embeddings.
2. A horizon-token formulation for anytime recurrence forecasting from partial blanking-period histories.
3. Internal evaluation on the DECAAF-II complete-record cohort and a larger auxiliary cohort without usable ablation geometry, reported separately from the full-model evaluation.

## 2 Related Work

**World models and medical world models.** World models learn state transitions for forecasting, simulation, and planning [3,9]. Predictive representation learning emphasizes compact latent dynamics that avoid brittle voxel-level reconstruction [2,4]. In medicine, generative modeling has increasingly shifted from visual realism toward clinically useful synthesis, translation, and task-aligned generation [26,6]. Recent medical world models include treatment-conditioned rollouts [19], diffusion-based post-treatment synthesis [23], context-aware latent transitions [8], and long-horizon electronic health record (EHR) patient simulation [15]. Related latent-transition ideas also appear in interactive imaging and radiology-focused predictive representation learning [10,24,25]. In contrast,



**Fig. 2. Overview of the proposed intervention-aware clinical world model.** Pre-ablation LGE-MRI is encoded into a spatial latent state  $z_0$ . Static covariates, the ablation heatmap, and time-ordered clinical events with ECGFounder embeddings condition ablation-conditioned initialization, latent updates, and elapsed-time drift. The final latent  $\hat{z}_T$  is matched to the training-only follow-up latent  $z_{\text{post}} = E_\theta(x_{\text{post}})$  and predicts recurrence probability and scar extent.

we focus on post-ablation structural trajectory forecasting by coupling 3D LGE-MRI, procedural geometry, irregular blanking-period events, and peri-event ECG embeddings.

**Atrial LGE-MRI and ablation outcome modeling.** Prior AF ablation studies segment and quantify the left atrium (LA) and scar [13,11]. Other work combines imaging and clinical features to predict recurrence [22,1,21,20]. Other studies [16,17,18] examine patient-specific and reinforcement-learning approaches to ablation strategy. SOFA [7] introduces action-conditioned scar simulation from ablation patterns. Unlike static recurrence predictors or scar simulators, our work models the post-procedural blanking period as an irregular event sequence and updates long-horizon recurrence risk by evolving a latent anatomical state.

### 3 Method

**Problem setup and targets.** The pre-ablation MRI  $x_0$  initializes the state; the follow-up MRI  $x_{\text{post}}$  is used only in training to define  $z_{\text{post}} = E_\theta(x_{\text{post}})$ . A terminal horizon token sets the query time  $t_T$ ;  $T$  is the forecast horizon, not the follow-up scan time. We forecast 451-day AF recurrence  $y$  and scar extent  $b$  using observations available by that horizon. Scar extent is the percentage of scar-

positive voxels in the follow-up LA wall mask; MAE is reported in percentage points. Figs. 1 and 2 summarize the timeline and model.

**Latent anatomical state from LGE-MRI.** A frozen variational autoencoder (VAE)  $(E_\theta, D_\theta)$ , pretrained on a larger imaging cohort, maps each MRI to a 3D spatial latent state  $z = E_\theta(x) \in \mathbb{R}^{C \times d \times h \times w}$ . We set the initial state to  $z_0 = E_\theta(x_0)$  and the training-only follow-up target to  $z_{\text{post}} = E_\theta(x_{\text{post}})$ .

**Conditioning signals and event tokens.** Static covariates are embedded as  $h_s = \phi_s(c_s)$ , and a 3D convolutional neural network (CNN) encodes the ablation heatmap as  $A = \phi_a(a) \in \mathbb{R}^{C_a \times d \times h \times w}$ . Each event at time  $t_i$  is mapped to a token

$$e_i = \left[ \frac{t_i - t_{\text{abl}}}{90}, \frac{t_T - t_{\text{abl}}}{180}, b_{\text{med}}, b_{\text{cv}}, b_{\text{rep}}, \psi_{\text{med}}(\text{med}_{\text{cat}}), u_i \right], \quad (1)$$

The first two scalar entries encode normalized event time and normalized queried horizon, respectively. The three binary flags indicate medication, cardioversion, or a repeat procedure. The function  $\psi_{\text{med}}$  embeds either a medication category or a dedicated **none** token. ECG context  $u_i$  mean-pools ECGFounder [12] embeddings recorded within 7 days before  $t_i$ . This window balances temporal relevance against sparse acquisition. Using only pre-event recordings avoids future information, and mean pooling handles variable recording counts. We set  $u_i = \mathbf{0}$  when no ECG is available.

We prepend an ablation anchor token at  $t_{\text{abl}}$ , sort clinical events by recorded time, censor them at day 90, and append the terminal horizon token at  $t_T$  with zero event flags. Variable-length sequences are padded and masked by  $m_i \in \{0, 1\}$ . A Transformer encoder produces contextualized embeddings

$$h_{1:L_{\text{max}}} = E_c(e_{1:L_{\text{max}}}, m_{1:L_{\text{max}}}), \quad (2)$$

For each token, we form  $c_i = \phi_c([h_s \| h_i])$ . These conditioning vectors are broadcast over the latent grid for the 3D convolutional transitions.

**Event-conditioned latent dynamics.** The first update ( $i=1$ ) uses the ablation anchor and encoded ablation map  $A$ . Later updates incorporate event context and elapsed-time drift. Starting from  $\hat{z}_0 = z_0$ , we update the latent state once per valid token  $i$ :

$$\Delta t_i = \frac{\max(t_i - t_{i-1}, 0)}{\tau}, \quad t_0 = t_{\text{abl}}, \quad (3)$$

$$\hat{z}_i = \hat{z}_{(i-1)} + f_\phi(\hat{z}_{(i-1)}, c_i, A) + \Delta t_i g_\phi(c_i), \quad (4)$$

Here,  $f_\phi$  is a 3D residual CNN, while the multilayer perceptron (MLP)  $g_\phi$  produces a context-dependent drift broadcast over space. They do not share parameters, and  $\tau=30$  days. The terminal token advances the state without adding a clinical event. To query a new horizon, we change only the time in this terminal token. Observed time gaps represent irregular sampling. Missing ECG is zero-filled, padding is masked, and unrecorded events are not imputed.

**Table 1.** Cohort construction and available modalities.

| Cohort                                            | N   | Use                             | Available data                                                                      |
|---------------------------------------------------|-----|---------------------------------|-------------------------------------------------------------------------------------|
| Parent union                                      | 830 | Availability audit              | Patients linked across available sources                                            |
| Complete full-modality cohort                     | 91  | Full-model evaluation           | Pre/post LGE-MRI, ablation geometry, blanking-period events, ECGs, scar, recurrence |
| Auxiliary cohort without usable ablation geometry | 258 | Evaluation without ablation map | Pre-MRI, static covariates, events, ECGs, recurrence; no usable ablation map        |
| Complete-cohort events                            | –   | Event schema                    | Medication 117, cardioversion 51, repeat procedure 3; median 2 events/patient       |

**Outcome heads and training objective.** We spatially average the predicted terminal latent  $\hat{z}_T$  and encoded ablation map  $A$  to obtain vector summaries  $\bar{z}$  and  $\bar{a}$ . These summaries are concatenated with static context for the recurrence and scar heads. We train  $(f_\phi, g_\phi)$  and the heads with the loss

$$\mathcal{L} = \lambda_z \|\hat{z}_T - z_{\text{post}}\|_2^2 + \lambda_{\text{cls}} \mathcal{L}_{\text{BCE}}(\hat{p}, y) + \lambda_{\text{bur}} \mathcal{L}_{\text{Huber}}(\hat{b}, b), \quad (5)$$

where  $\mathcal{L}_{\text{BCE}}$  is binary cross-entropy,  $\mathcal{L}_{\text{Huber}}$  is Huber loss, and  $z_{\text{post}}$  encodes the follow-up MRI. Inference excludes follow-up MRI; inputs are pre-ablation MRI, the ablation heatmap, static covariates, the event prefix, and available pre-event ECG embeddings. Later forecasts change only the terminal horizon token.

## 4 Experiments and Results

### 4.1 Datasets and Preprocessing

**Dataset and evaluation.** We use the DECAAF-II [14] cohort, a multicenter study with paired pre- and post-ablation LGE-MRI and procedural records that include ablation points. We pretrain  $E_\theta, D_\theta$  on a *disjoint* set of  $N=732$  patients with LGE-MRI but incomplete ablation or event records. The state-transition model is evaluated on  $N=91$  complete-record patients using 5-fold cross-validation with 3 seeds. Table 1 summarizes data availability and event frequency. For recurrence, we report the areas under the receiver operating characteristic and precision-recall curves (AUROC and AUPRC). For scar extent, we report MAE.

**Preprocessing.** Electroanatomic mapping (EAM) ablation points are manually mapped to MRI space and converted to a 3D heatmap. We resample both LGE-MRIs to 1.5 mm isotropic resolution, rigidly register the post-ablation scan to the pre-ablation scan, and transform the post-ablation masks accordingly. Volumes are cropped to the pre-ablation LA bounding box with a 10 mm margin and center-padded or cropped to (80, 80, 96). Intensities are normalized to  $[-1, 1]$  using robust region-of-interest (ROI) percentiles from the pre-ablation scan; the same parameters are applied to the registered post-ablation scan. Additional preprocessing details are provided in the code repository.

**Table 2.** Complete-record results ( $N = 91$ ; 5-fold CV, 3 seeds; mean $\pm$ SD). Post-MRI-only uses follow-up MRI at inference; sequence baselines use our allowed inputs.

| Method           | AUROC $\uparrow$                    | AUPRC $\uparrow$                    | Scar MAE $\downarrow$               |
|------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Static-only      | $0.511 \pm 0.126$                   | $0.551 \pm 0.070$                   | $6.507 \pm 1.667$                   |
| Post-MRI-only    | $0.525 \pm 0.045$                   | $0.575 \pm 0.079$                   | $3.189 \pm 0.698$                   |
| ECGFounder w/ECG | $0.639 \pm 0.134$                   | $0.635 \pm 0.133$                   | $4.773 \pm 1.403$                   |
| SOFA             | $0.610 \pm 0.142$                   | $0.660 \pm 0.134$                   | $3.510 \pm 0.903$                   |
| GRU              | $0.611 \pm 0.157$                   | $0.661 \pm 0.137$                   | $11.920 \pm 11.033$                 |
| LSTM             | $0.653 \pm 0.132$                   | $0.687 \pm 0.106$                   | $9.146 \pm 10.317$                  |
| Transformer      | $0.596 \pm 0.142$                   | $0.668 \pm 0.114$                   | $9.171 \pm 9.280$                   |
| <b>Ours</b>      | <b><math>0.756 \pm 0.051</math></b> | <b><math>0.777 \pm 0.058</math></b> | <b><math>2.971 \pm 0.675</math></b> |

**Table 3.** Auxiliary no-geometry results ( $N = 258$ ; 5-fold CV, 3 seeds; mean $\pm$ SD). Sequence baselines use the same inputs; this is not full-model validation.

| Method                       | Inputs                | AUROC $\uparrow$                    | AUPRC $\uparrow$                    |
|------------------------------|-----------------------|-------------------------------------|-------------------------------------|
| Static-only                  | Static                | $0.517 \pm 0.073$                   | $0.571 \pm 0.074$                   |
| Pre-MRI-only                 | MRI latent            | $0.529 \pm 0.081$                   | $0.594 \pm 0.078$                   |
| ECGFounder-only              | ECG                   | $0.596 \pm 0.039$                   | $0.608 \pm 0.038$                   |
| GRU                          | MRI+static+events+ECG | $0.555 \pm 0.059$                   | $0.592 \pm 0.063$                   |
| LSTM                         | MRI+static+events+ECG | $0.552 \pm 0.062$                   | $0.584 \pm 0.071$                   |
| Transformer                  | MRI+static+events+ECG | $0.549 \pm 0.072$                   | $0.581 \pm 0.069$                   |
| <b>Ours w/o ablation map</b> | MRI+static+events+ECG | <b><math>0.713 \pm 0.088</math></b> | <b><math>0.747 \pm 0.078</math></b> |

## 4.2 Results

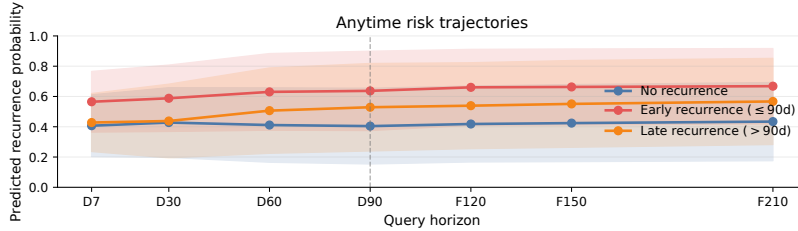
**Forecasting.** Using blanking-period information, we forecast diagnosis within 451 days post-ablation. In Table 2, Static-only and Post-MRI-only remain near chance, while ECGFounder and LSTM improve discrimination. Our 0.756/0.777 AUROC/AUPRC improves on matched-input LSTM by 0.103/0.090, indicating that state evolution, not input access alone, matters. OOF Brier score 0.201 and five-bin expected calibration error 0.032 show close coarse-bin probability alignment internally.

Fig. 3 shows that separation develops as blanking-period evidence accumulates: from D30 to D90, risk rises fastest for early recurrence, remains comparatively stable for no recurrence, and is intermediate for late recurrence. The ordering changes only gradually during F120–F210 no-new-event rollouts, indicating that the observed record establishes most of the separation. Overlapping SD bands, especially for late recurrence, nevertheless show substantial patient-level overlap.

Scar extent provides a structural check on the latent trajectory. Our method reaches MAE  $2.971 \pm 0.675$  percentage points versus  $3.189 \pm 0.698$  for oracle-style Post-MRI-only, despite not receiving follow-up MRI intensities at inference. The 0.218-point margin is small relative to fold variability, so the finding is competitive structural forecasting rather than reliable superiority over the oracle-style baseline. Relative to the cohort mean (interquartile range) of 8.61% (5.75–10.69%), the error remains nontrivial; no inter-observer benchmark is available.

**Table 4.** Component ablations on conditioning signals and model components.

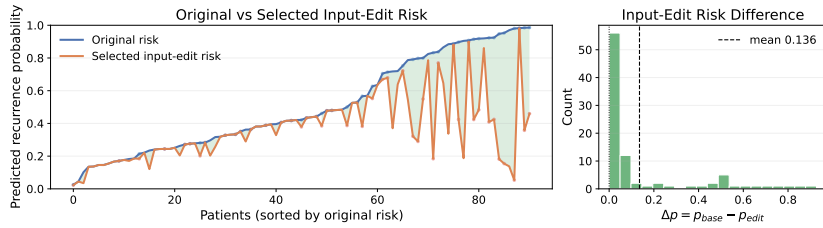
| Method                                     | AUROC $\uparrow$                    | AUPRC $\uparrow$                    |
|--------------------------------------------|-------------------------------------|-------------------------------------|
| <b>Ours (full)</b>                         | <b>0.756 <math>\pm</math> 0.051</b> | <b>0.777 <math>\pm</math> 0.058</b> |
| <i>Conditioning ablations</i>              |                                     |                                     |
| w/o terminal token                         | 0.742 $\pm$ 0.064                   | 0.738 $\pm$ 0.070                   |
| w/o AF ablation map                        | 0.711 $\pm$ 0.058                   | 0.758 $\pm$ 0.086                   |
| w/o ECGFounder                             | 0.705 $\pm$ 0.136                   | 0.748 $\pm$ 0.131                   |
| w/o events                                 | 0.623 $\pm$ 0.121                   | 0.652 $\pm$ 0.111                   |
| <i>Objective / architecture ablations</i>  |                                     |                                     |
| w/o latent matching loss ( $\lambda_z=0$ ) | 0.624 $\pm$ 0.200                   | 0.686 $\pm$ 0.173                   |
| Direct fusion (no step-wise evolution)     | 0.709 $\pm$ 0.095                   | 0.720 $\pm$ 0.077                   |

**Fig. 3. Anytime risk trajectories.** Mean $\pm$ SD for early recurrence ( $\leq 90d$ ,  $n=32$ ), late recurrence ( $>90d$ ,  $n=14$ ), and no recurrence ( $n=45$ ). D: observed-day; F: no-new-event horizons.

Because complete multimodal records with usable ablation geometry are scarce, we use  $N = 258$  additional patients to test whether the common-input rollout signal extends beyond the  $N = 91$  cohort. Without an ablation map, our variant obtains AUROC/AUPRC 0.713/0.747, gains of 0.158/0.155 over the strongest matched-input sequence baseline (Table 3). These values resemble the  $N = 91$  no-map ablation (0.711/0.758), although the cohorts are not directly comparable. The consistency indicates that event-conditioned rollout retains signal without geometry, while leaving the full model’s added spatial value untested in the larger cohort.

**Ablations.** Table 4 reveals a contribution hierarchy: removing events or latent matching produces the largest AUROC reductions (0.133 and 0.132), while removing the ablation map or ECG has smaller effects. Without latent matching, AUROC SD also rises from 0.051 to 0.200, consistent with structural supervision stabilizing the learned state. Direct fusion loses 0.047 AUROC and 0.057 AUPRC, supporting step-wise evolution over one-shot concatenation.

Shuffling event content while preserving times reduces AUROC/AUPRC from 0.756/0.777 to 0.709/0.713. Because event times and counts remain fixed, the drop indicates that event identity contributes beyond elapsed time.



**Fig. 4. Input-editing sensitivity.** **Left:** patient-wise  $\Delta p = p_{\text{base}} - p_{\text{edit}}$  after predefined edits. **Right:** cohort distribution of  $\Delta p$ . Edits are associational probes, not causal treatment effects.

Fig. 4 gives a second view of how the model uses blanking-period evidence. Specifically, we edit the presence or timing of medication, cardioversion, and repeat-procedure events and recompute recurrence risk, testing the model’s ability to compare alternative blanking-period scenarios. Edited and observed risks remain close for most patients, while a smaller subset with higher baseline risk shows large downward shifts. The resulting right-skewed  $\Delta p$  distribution has mean 0.136, but this value is subset-driven rather than a uniform cohort shift. The fitted model’s input dependence is therefore heterogeneous and not captured by the mean alone.

### 4.3 Limitations

The complete-record cohort is small ( $N=91$ ), and all results are internal to DECAAF-II. The  $N=258$  cohort lacks ablation geometry and cannot replace external validation. Manual EAM-to-MRI mapping can introduce localization error, while EAM measurements themselves may be noisy. Because clinical events may be missing or confounded, input edits show model sensitivity rather than causal effects. Alternative ECG windows, the effect of missingness patterns, latent-loss effects on scar MAE, and inter-observer scar variability remain unevaluated.

## 5 Conclusion

We presented an intervention-aware clinical world model that evolves a structured 3D latent state from ablation geometry, irregular blanking-period events, and ECG embeddings. In internal DECAAF-II cross-validation, it improves recurrence discrimination over static and matched-input sequence baselines, predicts scar extent without follow-up MRI at inference, and supports multi-horizon risk updates and retrospective input edits. Larger full-modality cohorts and external validation are needed before extending the framework toward causal modeling of post-ablation interventions.

## References

1. Atta-Fosu, T., LaBarbera, M., Ghose, S., Schoenhagen, P., Saliba, W., Tchou, P.J., Lindsay, B.D., Desai, M.Y., Kwon, D., Chung, M.K., et al.: A new machine learning approach for predicting likelihood of recurrence following ablation for atrial fibrillation from ct. *BMC Medical Imaging* **21**(1), 45 (2021)
2. Bardes, A., Garrido, Q., Ponce, J., Chen, X., Rabbat, M., LeCun, Y., Assran, M., Ballas, N.: Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471* (2024)
3. Bruce, J., Dennis, M.D., Edwards, A., Parker-Holder, J., Shi, Y., Hughes, E., Lai, M., Mavalankar, A., Steigerwald, R., Apps, C., et al.: Genie: Generative interactive environments. In: *Forty-first International Conference on Machine Learning* (2024)
4. Burchi, M., Timofte, R.: Mudreamer: Learning predictive world models without reconstruction. *arXiv preprint arXiv:2405.15083* (2024)
5. Calkins, H., Hindricks, G., Cappato, R., Kim, Y.H., Saad, E.B., Aguinaga, L., Akar, J.G., Badhwar, V., Brugada, J., Camm, J., et al.: 2017 hrs/ehra/ecas/aphrs/solaece expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Ep Europace* **20**(1), e1–e160 (2018)
6. Chung, Y., Darzi, A.E., Khoury, C.E., Feng, H., Marrouche, N., Hamm, J.: Craft: Clinical reward-aligned finetuning for medical image synthesis. *arXiv preprint arXiv:2605.12650* (2026)
7. Chung, Y., Lim, C., Bidaoui, G., Massad, C., Marrouche, N., Hamm, J.: Sofa: Deep learning framework for simulating and optimizing atrial fibrillation ablation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 500–509. Springer (2025)
8. Ding, T., Zou, Y., Chen, C., Shah, M., Tian, Y.: Clarity: Medical world model for guiding treatment decisions by modeling context-aware disease trajectories in latent space. *arXiv preprint arXiv:2512.08029* (2025)
9. Hafner, D., Pasukonis, J., Ba, J., Lillicrap, T.: Mastering diverse control tasks through world models. *Nature* pp. 1–7 (2025)
10. Jiang, H., Sun, Z., Jia, N., Li, M., Sun, Y., Luo, S., Song, S., Huang, G.: Cardiac copilot: Automatic probe guidance for echocardiography with world model. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 190–199. Springer (2024)
11. Lefebvre, A.L., Yamamoto, C.A., Shade, J.K., Bradley, R.P., Yu, R.A., Ali, R.L., Popescu, D.M., Prakosa, A., Kholmovski, E.G., Trayanova, N.A.: Lassnet: a four steps deep neural network for left atrial segmentation and scar quantification. In: *Challenge on Left Atrial and Scar Quantification and Segmentation*, pp. 1–15. Springer (2022)
12. Li, J., Aguirre, A., Moura, J., Liu, C., Zhong, L., Sun, C., Clifford, G., Westover, B., Hong, S.: An electrocardiogram foundation model built on over 10 million recordings with external evaluation across multiple domains. *arXiv preprint arXiv:2410.04133* (2024)
13. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: Atrialjsqnet: a new framework for joint segmentation and quantification of left atrium and scars incorporating spatial and shape information. *Medical image analysis* **76**, 102303 (2022)
14. Marrouche, N.F., Wazni, O., McGann, C., Greene, T., Dean, J.M., Dagher, L., Kholmovski, E., Mansour, M., Marchlinski, F., Wilber, D., et al.: Effect of mri-guided fibrosis ablation vs conventional catheter ablation on atrial arrhythmia recurrence in patients with persistent atrial fibrillation: the decaaf ii randomized clinical trial. *Jama* **327**(23), 2296–2305 (2022)

15. Mu, L., Huang, Z., Gu, Y., Qin, S., Zhang, S., Zhang, X.: Ehrworld: A patient-centric medical world model for long-horizon clinical trajectories. *arXiv preprint arXiv:2602.03569* (2026)
16. Muffoletto, M., Qureshi, A., Zeidan, A., Muizniece, L., Fu, X., Zhao, J., Roy, A., Bates, P.A., Aslanidi, O.: Toward patient-specific prediction of ablation strategies for atrial fibrillation using deep learning. *Frontiers in Physiology* **12**, 674106 (2021)
17. Muizniece, L., Bertagnoli, A., Qureshi, A., Zeidan, A., Roy, A., Muffoletto, M., Aslanidi, O.: Reinforcement learning to improve image-guidance of ablation therapy for atrial fibrillation. *Frontiers in Physiology* **12**, 733139 (2021)
18. Ogbomo-Harmitt, S., Muffoletto, M., Zeidan, A., Qureshi, A., King, A.P., Aslanidi, O.: Exploring interpretability in deep learning prediction of successful ablation therapy for atrial fibrillation. *Frontiers in Physiology* **14**, 1054401 (2023)
19. Qazi, M.A., Nadeem, M., Yaqub, M.: Beyond generative ai: World models for clinical prediction, counterfactuals, and planning. *arXiv preprint arXiv:2511.16333* (2025)
20. Razeghi, O., Kapoor, R., Alhusseini, M.I., Fazal, M., Tang, S., Roney, C.H., Rogers, A.J., Lee, A., Wang, P.J., Clopton, P., et al.: Atrial fibrillation ablation outcome prediction with a machine learning fusion framework incorporating cardiac computed tomography. *Journal of cardiovascular electrophysiology* **34**(5), 1164–1174 (2023)
21. Roney, C.H., Sim, I., Yu, J., Beach, M., Mehta, A., Alonso Solis-Lemus, J., Kotadia, I., Whitaker, J., Corrado, C., Razeghi, O., et al.: Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific left atrial models. *Circulation: Arrhythmia and Electrophysiology* **15**(2), e010253 (2022)
22. Varela, M., Bisbal, F., Zacur, E., Berruezo, A., Aslanidi, O.V., Mont, L., Lamata, P.: Novel computational analysis of left atrial anatomy improves prediction of atrial fibrillation recurrence after ablation. *Frontiers in physiology* **8**, 68 (2017)
23. Yang, Y., Wang, Z.Y., Liu, Q., Sun, S., Wang, K., Chellappa, R., Zhou, Z., Yuille, A., Zhu, L., Zhang, Y.D., et al.: Medical world model: Generative simulation of tumor evolution for treatment planning. *arXiv preprint arXiv:2506.02327* (2025)
24. Yue, Y., Wang, Y., Jiang, H., Liu, P., Song, S., Huang, G.: Echoworld: Learning motion-aware world models for echocardiography probe guidance. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 25993–26003 (2025)
25. Yue, Y., Wang, Y., Tao, C., Liu, P., Song, S., Huang, G.: Chexworld: Exploring image world modeling for radiograph representation learning. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 20778–20788 (2025)
26. Zhou, X., Li, C., Wang, S., Li, Y., Tan, T., Zheng, H., Wang, S.: Generative artificial intelligence in medical imaging: Foundations, progress, and clinical translation. *arXiv preprint arXiv:2508.09177* (2025)