We thank the reviewers for their constructive feedback on our paper. We will take this feedback into account when revising the paper. Below we provide point-by-point response.

Rev.#1:

Reg. useless intervals and method breakdown: When $y_\alpha(x, \mathbf{z}) = \max(\mathcal{Y})$, the upper limit is noninformative which occurs when $(x, \mathbf{z})$ is 'far' from the training data. In such cases, however, the proposed policy would fall back onto the logging policy if $p(\mathbf{z}|x)$ is modelled accurately and used in the weights $w(x, \mathbf{z})$ in (8). The method would indeed perform poorly if the model were highly inaccurate, which would result in weights that may alter the decisions drastically. Fortunately, it is possible to use model validation techniques to assess the accuracy of the model, as pointed out in Sec. 3.2. Even in the case of misspecified models, the weights can be sufficiently accurate to provide accurate intervals, as Figure 3d demonstrates for the synthetic example in Sec. 4.1, where the empirical estimate of $\mathbb{P}^{\pi_\alpha}(y \leq \ y_\alpha(\mathbf{z}))$ matches the theoretical limit well.

Reg. motivation for locally weighted average $\mu(x, y, \mathbf{z})$: Firstly, it leads to computationally efficient conformal limits since $\mu(x, y, \mathbf{z})$ must be fitted for each evaluation at $(x, y, \mathbf{z})$. Using the nonparametric weighted average, each fitting has a constant runtime $\mathcal{O}(1)$. Secondly, using parametric predictive models $\tilde{\mu}(x, y, \mathbf{z})$ yields conformal limits that are more sensitive to model misspecification. Indeed, misspecified parametric models may produce larger residuals and hence larger conformal limits than a nonparametric locally weighted average. These points will be highlighted in the revised manuscript.

Reg. complexity of Algorithm 1: The main operations which depend on the number of datapoints n are lines 3, 4 and 10. Out of these, line 10 involves a sorting operation to compute the quantile $s_{1-\alpha}(\widehat{F})$ which dominates all other computations and results in a total runtime $\mathcal{O}(n \log\ n)$. The method thus is scalable to large n.

Reg. comparison to reasonable baselines: We have now added another baseline which explicitly learns a policy using an consistent estimate of α-quantile of the costs y (based on the cited paper by Wang et. al.). For the synthetic case in Sec. 4.1, it results in a slightly lower α-quantile level, but significantly higher tail costs beyond the α-quantile as compared to the proposed method. We will include comparisons with this additional baseline for both numerical examples in the revised manuscript or supplementary material.

Reg. discussion of relevant work in reinforcement learning: We will add the suggested references and additional references on safety-critical applications.

We will add a clarification on the notation $\mathbb{P}^x$.

Rev.#2:

Reg. unconfoundedness: We agree with the reviewer and we do assume that there are no unobserved confounders. We will make this assumption explicit in the revised manuscript.

Reg. overlap: Result 1 does require overlap $p(x|\mathbf{z}) > 0$ in order for the weights $w(x, \mathbf{z})$ in eq. (8) to be finite. However, as pointed out at the end of Sec. 3.1, as the evaluated weight $w(x, \mathbf{z}) \to \infty$, then $p_x(x, \mathbf{z}) \to 1$ and the conformal limit simply becomes uninformative so that the method remains operational even for infinite weight $w(x, \mathbf{z})$. We will clarify this point in the revised paper.

Rev.#3:

Reg. distributional shift: If the feature training distribution $p(\mathbf{z})$ shifts from the test distribution, say, $q(\mathbf{z})$, then the method can be readily extended to compensate for such distributional shifts, provided that $q(\mathbf{z})$ can be evaluated at a given $\mathbf{z}$. We will include this remark in the revised manuscript.

Reg. dimension reduction and high-dimensional data: We have not studied the effect of dimension reduction on the performance of our method. However, it is possible to check the accuracy of the learned generative model $\widehat{p}(\mathbf{z}|x\ =\ k)$ using the model validation methods referred to in Sec. 4.2. This provides a guideline for choosing the appropriate feature dimension to which data is can be reduced.

Reg. validity of results under confounding: Indeed, we do assume no unobserved confounders and will make the assumption explicit.

Reg. fairness: We have not explored this question in this work but included a remark on it in the broader impact section.

Rev.#4:

Reg. comparison to other methods: We do compare our method to *mean-optimal* policy $\pi(\mathbf{z})$, which is a standard method considered in the literature. Moreover, we have also included an additional baseline as explained in the reply to Reviewer 1.