The generative approach assumes, it is possible to capture a latent state z and use this state to generate suitable counterfactual candidates. We condition the generation procedure on the factual instance to generate counterfactuals that show sparse differences to the original sequence. The core idea is to sample randomly $e^{*} \sim p(z|e)$ to generate counterfactual candidates. We can sort each candidate by their *viability* and choose top-K contenders as viable couunterfactuals. There are a multitude of approaches to generate the counterfactuals. However, we will limit our exploration to a sequential Variational Autoencoders (VAEs). VAEs approximate $p(z|e)$ by trying to reconstruct the input using Monte-Carlo methods. For this purpose we encode a sequence into a mean vector. Then we use this vector to reconstruct the initial input again. The architecture resembles the predictor model's architecture, as you can see in Figure 1.

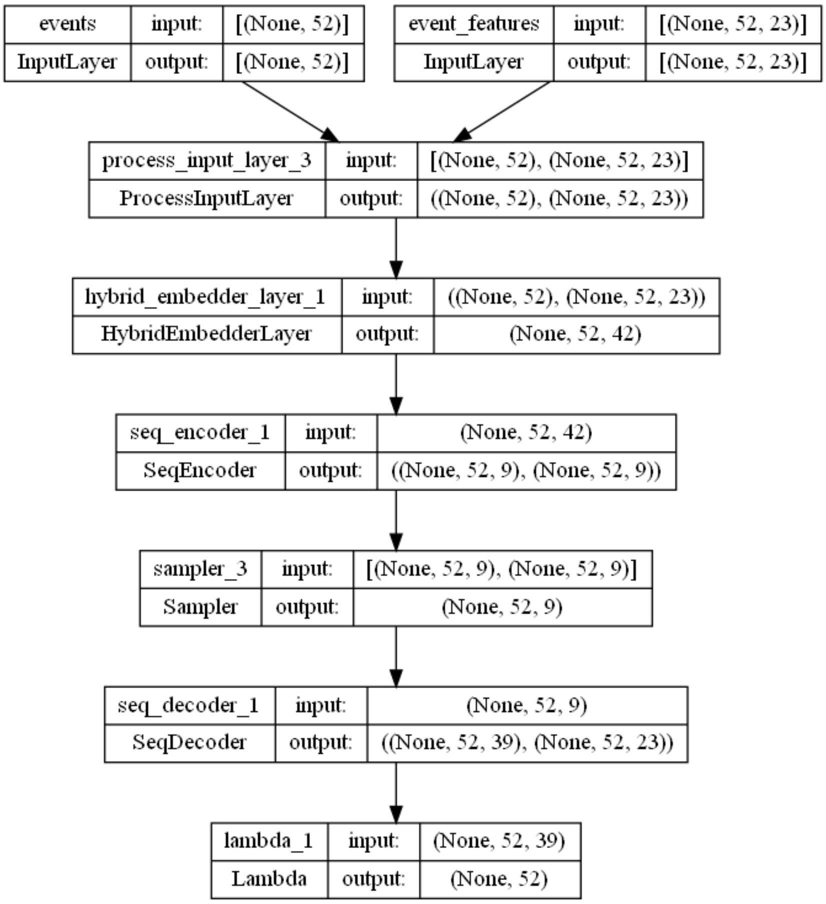

Figure 1: Shows the different components of the VAE architecture. Each elements contains information about the input and output of a layer. None is a placeholder for the batch size.