To generate counterfactuals, we will need to establish a conceptual framework, which consists of three main components. The three components are shown in Figure 1.
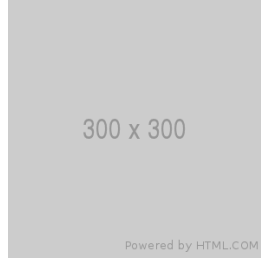

Figure 1: Shows a visual representation of the methodological framework.

The first component is a predictive model. As we attempt to explain model decisions with counterfactuals, the model needs to be pretrained. We can use any model that can predict the probability of a sequence. This condition holds for models trained for process outcome classification and next-activity prediction. The model used in this thesis is a simple LSTM model using the process log as an input. The model is trained to predict the next action given a sequence.

The second component is a generative model. The generative model produces counterfactuals given a factual sequence. In our approach, each generative model should be able to generate a set of counterfactual candidates given one factual sequence. Hence, we cannot use purely deterministic generators. Therefore, we compare two different generative approaches. First, a sequential Variational Autoencoder (VAE) and then an evolutionary approach. Both methods allow us to use a factual sequence as starting point for the generative production but also generate multiple variations of the final solution.

The generated candidates are subject to the third major component's scrutiny. To select the most *valid* counterfactual candidate, we evaluate their validity score using a custom metric. The metric incorporates all main criterions for valid counterfactuals mentioned in **??**. We measure the *sparcity* between two sequences using a multivariate sequence distance metric. The *precision* of the prediction will compare the likelihood of the counterfactual outcome without changes to the sequence and the counterfactual candidates. For this purpose, we require the predictive model to compute their likelihoods. We measure *diversity* by [a way to compute the diversity among counterfactuals]. Lastly, we need to determine the feasibility of a counterfactual. This requires splitting the feaibility in two parts. First,

the likelihood of the sequence of each event and second, the likelihood of the features given the event.