Hsieh, Moreira, and Ouyang follow a similar pattern of assessing the quality of their counterfactuals. The authors also focus on similarity, sparsity, feasibility and likelihood-improvement. However, they incorporate and operationalise them differently. Their approach is most apparent in their loss function.

Similarity: Similar to our approach, the authors use a distance function and optimise it using gradient descent. They evaluate the quality of their counterfactuals using the same function[1]. However, we use a modified Damerau-Levenshstein distance algorithm to incorporate structural differences such as the sequence lengths or transposed events.

Sparsity: The DiCE4EL approach does not consider this.

Feasibility: This quality criterion is embodied by two loss functions: Category loss and scenario loss. The category loss ensures that categorical variables remain categorical after generation. The scenario loss adds emphasis on only generating counterfactuals that are in the event log. Unlike our probabilistic interpretation, they treat the existence of feasible counterfactuals as a binary criterion[2].

Likelihood: Similar to the authors' scenario loss, they treat the improvement of a class as a binary state. Either the counterfactual changes the model's prediction of the desired outcome, or it does not.

The details of each criterion's operationalisation are explained in [1]. By assessing their interpretation of quality criteria, we see the clear distinction between our approach and the approach of Hsieh, Moreira, and Ouyang.

First, their viability measure decisively discourages the generation of counterfactuals that are not present in the dataset. In contrast, our approach treats this aspect as a soft constraint.

Second, while our approach acknowledges general improvements in likelihoods, DiCE4EL treats all counterfactuals that do not lead to better desires as detrimental solutions. However, one can argue that improving the likelihood of the desired outcome just slightly is already beneficial.

Third, Hsieh, Moreira, and Ouyang do not optimise sparsity, while we include it within our framework. One can argue that similarity automatically incorporates aspects of sparsity, but we disagree with this notion. We can see this by employing a simple example: Let factual A have features signifying the biological sex (binary), the income (normalised) and the age (normalised)

[1] They call it proximity during evaluation

[2] They call it plausibility during evaluation

$\begin{pmatrix}1\\1\\1\end{pmatrix}$ as event attributes. Let counterfactual B have the same event attributes with $\begin{pmatrix}0\\1\\1\end{pmatrix}$. Let's assume the distance measure uses the L1-norm. Then, a counterfactual C with event attributes $\begin{pmatrix}1\\0.5\\0.5\end{pmatrix}$ would have the same distance to factual A as B has. However, C requires the change of two event attributes, while B only requires 1 change. In a scenario in which we seek to reduce the number of edits, B is preferable to C, regardless of the distance to A.

The last difference stems from the fact that Hsieh, Moreira, and Ouyang do not include structural sequence characteristics in their similarity measure.