Table 1 shows how each model scores under different operationalisations of viability aspects. They were derived from Hsieh et al.'s custom evaluation protocol and aim to provide a better comparison. Each value reflects the mean across all counterfactual results per model.

Table 1: Shows the mean result of each models' result with respect to diversity, plausibility proximity and sparsity.

	Property	Diversity	Plausibility	Proximity	Sparsity
Model	Dimension				
Casebased Generator	Activity	0.007850	1.000000	12.545000	9.345000
	Resource	0.006100	0.000000	15.710000	15.505000
Evoluationary: SBI-ES-OPC-SBM-FSR	Activity	0.375000	0.000000	15.250000	13.250000
	Resource	0.250000	0.000000	15.750000	15.750000
Random Generator	Activity	0.005000	0.000000	23.415000	21.160000
	Resource	0.193300	0.000000	24.185000	24.185000

The results show that diversity is the highest for the evolutionary algorithm in terms of activity traces and resource traces. The Random-Search Generatordisplays low diversity for activities generated and a higher diversity for the resource.

Only the Casebased-Search Generatorreaches a maximum score of 1 for plausibility. All the other models are far below or 0.

In terms of proximiny, the Casebased-Search Generatorhas the lowest actvity prximity. The average distance is 12.55. The SBI-ES-OPC-SBM-FSR Generatortakes the second place. Interestingly, the gap between the proximity for activities is larger than the gap between proximities in terms of resources.

Again, the Casebased-Search Generatorhas the lowest sparcity with 9.34 in terms of activity but only remains slightly better than SBI-ES-OPC-SBM-FSR Generatorin terms of resources.