

Peer Review

Review of: "Grade Inflation in Generative Models"

Taseef Rahman¹

1. George Mason University, United States

This paper addresses a very interesting topic that can be trivial to miss - do we have the correct evaluation of data generated by generative models? At first glance, the answer might seem obvious given that we have a myriad of ways to measure the correctness of the synthetic data with respect to the original dataset. But, as demonstrated by the authors, the answer might not be so obvious. Here are my thoughts - 1. The approach of determining the annulus score directly corresponds to how we as humans tend to distinguish between two data distributions. While novel, it would benefit from a larger sample of human ratings as it would have decreased vulnerability to sampling error. 2. One of the approaches that the authors could have undertaken to address the higher dimensionality is to reduce the initial dimensionality followed by cross-validation on a sample downstream task. One such sample task could be the determination of document similarity. This could be facilitated by text embedding similarity followed by human evaluation as utilized by the authors for their approach.

Declarations

Potential competing interests: No potential competing interests to declare.