

Examining Diffusion Model Architectures for Text Generation: Challenges and Autoregressive Comparison

Diffusion models are currently state-of-the-art (SOTA) in generating images and audio, but they still lag behind autoregressive models (AR-LM) for text generation. A key drawback of autoregression is that if an error occurs at any step, it propagates through the following tokens, and these models are resource-intensive since they generate text one token at a time. Diffusion models offer the advantage of controllability through auxiliary models, but applying them to text is more challenging due to issues with data representation and compatibility with existing classifiers.

