# Model Limitations

1. Limited Dataset Size: The model was trained on a relatively small dataset, limiting its ability to generalize to broader contexts or varied data.

2. Imperfect Masking: The masking of entities like phone numbers, emails, and URLs is not always perfect. Some complex patterns may not be accurately identified.

3. Bias in Pre-trained Model: Using pre-trained models like BERT may introduce inherent biases present in the training data, affecting the neutrality of the results.

4. Contextual Limitation: The model may struggle with understanding context when masking entities, especially in ambiguous sentences.

5. Memory and Computation Limitation: Fine-tuning BERT or Huggingface models requires high computational power and sufficient GPU memory, which may not always be available.

6. Overfitting: Due to limited data and fine-tuning on a small dataset, the model might have overfitted to the training data, impacting its performance on unseen data.

7. False Positive Rate: The model sometimes identifies non-entity words as entities (false positives), reducing the accuracy of masking.

8. Scalability: Deploying the model on larger datasets or in production environments may require further optimization to reduce latency and resource consumption.

9. Limited Entity Types: The model is currently focused on masking phone numbers, emails, and URLs. Expanding it to cover other PII (Personally Identifiable Information) will require further training.

10. Data Privacy: Although the model masks sensitive information, storing or processing unmasked data may still pose data privacy concerns.