

Supplementary Materials for Enabling Text-free Inference in Language-guided Segmentation of Chest X-rays via Self-guidance

Lesion Localization and Report Generation

The model's precision in identifying infected areas and the quality of the generated medical reports were further evaluated (refer to Tables S1 and S2). With self-localization accuracy at 76.2%, recall at 81.2%, and precision at 80.2%, the model proves highly effective in accurately identifying infected areas. The high scores in BLEU, METEOR, and ROUGE-L metrics affirm that the generated reports are closely aligned with the ground truth in terms of content and expression. This efficiency in accurately locating infections and generating coherent reports explains why our method achieves performance comparable to that with ground truth input, relying solely on images during inference.

Table S1. Performance evaluation for lesion localization

Subset	Accuracy	Recall	Precision	F-Score
Train	0.984	0.987	0.986	0.985
Validation	0.728	0.770	0.713	0.718
Test	0.762	0.812	0.802	0.795

Table S2. Performance evaluation for report generation

Subset	Bleu 1	Bleu 2	Bleu 3	Bleu 4	Meteor	Rouge-L
Train	0.839	0.826	0.813	0.797	0.886	0.805
Validation	0.808	0.732	0.661	0.601	0.879	0.804
Test	0.860	0.797	0.733	0.677	0.902	0.848