Methods:

We propose Medical Vision Language Learner (MedViLL), a multi-modal pre-training model for medical images and reports with a novel self-attention scheme. 2) We demonstrate the effectiveness of our approach with detailed ablation study on extensive vision-language understanding and generation-based downstream tasks, including diagnosis classification, medical image-report retrieval, medical visual question answering, and radiology report generation. 3) We demonstrate the generalization ability of our approach under the transfer learning setting using two separate Chest X-ray datasets, where we pre-train a model on one dataset and perform diverse downstream tasks on another.

Techniques:

Diagnosis classification

Medical Image report retrieval

Medical visual question answering

Radiology report generation

Key Concepts in AI Safety: An Overview