

Information of Submitted Datasets

The original data used for data augmentation (i2b2 2006, i2b2 2014) requires permission for use, so it cannot be disclosed to third parties without consent. Since the augmented data is derived from the original data, its disclosure is also restricted. However, the zero-shot prompt-generated data in our paper does not use the original data, so there is no issue with the data use agreement, and we intend to disclose only the zero-shot augmented data. For detailed information regarding the data use agreement, please refer to [Harvard Medical School Department of Biomedical Informatics](#). (You can view the terms when applying to download the data.)

The detailed process of augmenting the data by repeating augmentation levels alpha 1 to 5 five times, accumulating the data at each level, and then splitting it into training and validation sets can be found in section 4.4 of the paper. The test set, being part of the i2b2 2006 and i2b2 2014 data, cannot be disclosed.