

Transformer-Based Polymer Property Prediction Model

STAGE 1: MLM Dataset (CSV format)

STAGE 2: Regression Dataset (TXT format)

STAGE 1

INPUT LAYER

Polymer SMILES Sequence
Example: "CC(C)(C)C1=CC=C(C=C1)C(=O)O"

RoBERTa-style Tokenization

Tokenized Sequence + Position Embeddings

TRANSFORMER ENCODER

Multi-Head Self-Attention Layers

Position-wise Feed-Forward Networks

Contextual Representations

STAGE 1

MLM PRETRAINING HEAD

Linear + Activation Layer

Cross-Entropy Loss

RECONSTRUCTED SMILES SEQUENCE

PROPERTY REGRESSION HEAD

MLP Regressor Layers

Mean Squared Error Loss

PREDICTED POLYMER PROPERTIES

STAGE 2