TRANSPOLYMER

INTRODUCTION:

TransPolymer is a Transformer-based language model designed for polymer property prediction by using SMILES tokenizer and Masked Language Modeling (MLM) pretraining to enhance accuracy and efficiency.

SCOPE OF PROJECT:

TransPolymer uses AI to revolutionize polymer research , making it easier , faster , and more accurate to predict how different polymers will behave .

USERS:

◆ Material Scientists ◆ Polymer Chemists ◆ Computational Chemists ◆ R&D Engineers
◆ Data Scientist in Material Informatics ◆ Input and Output

Pain Points:

◆ Needs high-quality data
◆ External factors (temperature, aging) not always captured
◆ High computational power required
◆ Struggles with unseen polymers

01

TECH STACK:

◆ MERN Stack
◆ Python
◆ PyTorch
◆ Hugging Face Transformers
◆ RDKit and Scikit-learn
◆ NumPy, Pandas and Matplotlib.

02

Business Case:

◆ Cuts Costs & Saves Time
◆ Speeds Up Material Discovery
◆ Better Decisions with Data
◆ Gives Companies an Edge

03

04

Roles Involved:

◆ Researcher/User
◆ End Users(Polymer Engineers/Chemists)

Dataset and Data Source:

We created a custom dataset which contains 6 polymer properties. The dataset was created by referring the information extracted from GitHub repositories related to polymer science.

Cloud Deployment:

◆ Train & Store → Train on Cloud GPUs, store in S3/GCS.
◆ Deploy API → Use FastAPI/Flask, deploy via Docker (K8s, Lambda, Cloud Run).
◆ Frontend → React/Vue + JWT auth for secure access.
◆ Monitor & Scale → Prometheus + Grafana, auto-scale with K8s/Cloud Run.

Data Preprocessing :

◆ Verification (Check if the Data is Correct)

◆ Validation (Ensure Data is Meaningful & Usable)

◆ Normalization (Standardizing Data for Better Model Performance)

TECH STACK

MERN Stack · Python · PyTorch · RDKit · Scikit-learn · Hugging Face Transformers · NumPy · Pandas · Matplotlib

CLOUD SERVICES

Model Monitoring · Datasets & Model Checkpoints · Model Training · API Deployment

train model · track performance · store datasets · deploy API

BACKEND

Request Handling → tokenize SMILES & transform input → Preprocessing Layer → encode sequences → Transformer Model

store predictions → Database Operations → save data → Cloud Storage & Logging

DATABASE

Polymer Data & Predictions · Model Logs Collection · Users Collection · Model Versions & Hyperparameters

END USERS

Scientists & Chemists · AI & Software Engineers

interact with application

FRONTEND

UI → send requests → Fast API → JWT login → Authentication

SMILE'S String Input

Fast API → manage UI states & cache responses → State Management

""REST/GraphQL""