Supplementary information

Unified Multimodal Multidomain Polymer Representation for Property Prediction

Qi Huang[‡a, b, c], *Yedi Li*[‡a], *Lei Zhu*[a] *, *Qibin Zhao*[c] *, *Wenjie Yu*[a]

a National Key Laboratory of Materials for Integrated Circuits, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences.

b College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences.

c RIKEN Center for Advanced Intelligence Project (RIKEN AIP)

Email: leizhu@mail.sim.ac.cn, qibin.zhao@riken.jp

Keywords: Polymer Representation; Property Prediction; Polymer Caption; Large Language Model;

Prompt and Examples for Polymer Caption Generation

Motivation and Refinement Iterations of the Prompt

Our prompt engineering strategy was driven by three primary objectives: (1) grounding the LLM in domain-specific knowledge while minimizing hallucinations, (2) ensuring consistency in both style and content for caption generation to facilitate downstream property prediction tasks, and (3) balancing the richness of information with the prevention of data leakage. we performed four iterative refinement cycles in close collaboration with polymer experts, systematically improving both the design and performance of the prompts.

Initial Attempts. Our initial prompts were intentionally simple—for example, "Generate a caption for the following polymer with SMILES…" or "Provide a caption for the polymer with the given SMILES representation...". However, the outputs were often overly simplistic and lacked critical domain-specific details(only in one sentence), indicating the need for more sophisticated guidance.

Refinement through Role Definition. To address this, we explored more structured prompt engineering strategies. A key insight during this phase was the importance of role specification and task definition[1]. Through several experiments, we modified the prompt to guide the model to assume the identity of "an expert in polymer science", which significantly enhanced the technical rigor and relevance of the outputs. Additionally, with input from domain experts, we imposed content and stylistic constraints on the captions, integrating these refinements into our second prompt version. Additionally, we incorporated content and stylistic constraints, developed in consultation with domain experts, into our second prompt version. While this improved the level of detail, it also introduced factual inaccuracies. For instance, even for well-known polymers like polystyrene, the model occasionally failed to correctly infer the monomer structure or polymer name from the SMILES input.

Integration of knowledge-enhanced Inputs. In the third refinement cycle, we focused on improving factual accuracy using a knowledge-enhanced prompting strategy inspired by Retrieval-Augmented Generation (RAG). Classical RAG operates in two sequential stages: retrieving relevant information from the database, followed by generating text based on the retrieved content. Given that all relevant polymer information was already available, we bypassed the retrieval step and directly embedded key contextual information into the prompt.

Through expert consultation and iterative evaluation, we selected essential polymer-related information while exclude detailed polymer properties to prevent data leakage. Two alternative input formats were tested for integrating this information: (1) unstructured text, where polymer details were embedded in a fixed-template paragraph, and (2) structured data, formatted as JSON. Empirical comparisons showed no significant difference in output quality, highlighting the LLM's robust capability in comprehending structured data. However, he JSON-based format proved more efficient in terms of token usage and preprocessing, making it the preferred choice.

This refinement effectively ensured that the generated captions contained accurate polymer information as provided. However, the generated text often exhibited excessive verbosity while lacking substantive insights. In some cases, the model simply reiterated JSON fields verbatim rather than synthesizing them into coherent, natural-language descriptions, likely due to its tendency to preserve structure faithfully.

Refining Output Structure through Few-Shot Examples. To address residual ambiguity in task expectations (e.g., tone, depth, and section structure), we introduced few-shot examples. We experimented with different quantities, finding that 3–4 examples provided the best balance: a single example led to overly rigid mimicking, while too many caused the model to overfit to unrelated phrasing. The final prompt included three manually curated examples (PP, PMMA, and PLA), which effectively guided the model toward generating factually accurate, stylistically consistent, and

comprehensive captions. These outputs reliably included four key sections: basic details, structure, properties, and applications.

Through this multi-stage refinement process, we developed a final prompt architecture comprising: (i) an explicit expert role specification (ii) a structured JSON input encoding essential polymer context, and (iii) a carefully curated set of few-shot examples. This approach enables the generation of high-quality, domain-specific captions that embed structured polymer knowledge in fluent and informative natural language.

Composition of the Prompt:

Our final prompt consists of the following three parts.

Role & Task Definition

You are an expert in polymer science, tasked with applying specialized knowledge to craft brief and insightful caption for polymer. Leveraging the structured data provided, your goal is to combine structured information with expert analysis to enhance the understanding of polymer structures and functions. The output caption should:

Basic Details: *Include the polymer's name, chemical class, molecular formula, and synthesis method.*

Structural Features: *Highlight unique structural aspects, such as specific functional groups or molecular arrangements.*

Key Properties: *Emphasize notable properties, such as mechanical strength, thermal stability, biocompatibility, or transparency.*

Applications: *Relate the polymer's properties to practical uses in industries or technologies.*

Polymer structured data

Here is the structured polymer JSON Data:

```
{
    "polymer_information": {
        "smiles_representation": "[*]CC([*])c1ccccc1",
        "polymer_class": ["Polystyrenes", "Polyvinyls"],
        "structure_name": "poly(1-phenylethylene)",
        "source_name": "polystyrene",
        "other_name": ["poly(vinylbenzene)", "poly(styrene)"],
        "abbreviation": "PS",
        "chemical_formula": "C8H8",
        "formula_weight": 104.15,
        "synthesis_reaction": {
            "monomer": "styrene",
            "reaction_type": "addition",
            "reaction_mechanism": "radical",
            "reaction_conditions": "70°C, N2 atmosphere",
        }
    }
}
```


Example Captions

Please refer to the following example and use clear, concise, and scientifically accurate language.

"Poly(1-methylethylene), or polypropylene (PP), is a lightweight, semi-crystalline polyolefin with the molecular formula C3H6 and a formula weight of 42.08 g/mol. Synthesized through the addition polymerization of propylene monomers, it features a unique isotactic arrangement of methyl groups that enhances crystallinity and strength. Known for its exceptional chemical resistance, thermal stability, and low density, PP is widely utilized in packaging, automotive parts, textiles, and household goods. Its durability, affordability, and versatility make it indispensable in applications requiring long-term performance and cost-effectiveness."

"Poly(methyl methacrylate) (PMMA) is a transparent, lightweight polymer belonging to the polyacrylic and polyvinyl classes, with the molecular formula C5H8O2 and a molecular weight of 100.12 g/mol. It is typically synthesized through radical addition polymerization in solution. PMMA features a rigid backbone structure with methyl

ester functional groups, contributing to its exceptional optical clarity, mechanical strength, and resistance to UV radiation and weathering. These properties make it indispensable for applications such as optical lenses, protective transparent barriers, signage, and lightweight glass alternatives, particularly in architectural, automotive, and medical industries."

"Poly(lactic acid) (PLA), a biodegradable aliphatic polyester with the molecular formular C3H4O2 and a molecular weight of 100.12 g/mol, is synthesized through the ring-opening polymerization of L-lactide. Distinguished by its renewable origin and biocompatibility, PLA features ester linkages that enable controlled biodegradation. It exhibits notable thermal processability and mechanical properties suitable for diverse applications. Widely utilized in biomedical fields for sutures, drug delivery systems, and tissue scaffolds, PLA's eco-friendly and thermoplastic nature also makes it an ideal material for sustainable packaging, disposable goods, and 3D printing."

Sample Output

Based on the aforementioned prompts, we used the GPT-4 API (gpt-4o-2024-08-06) to generate the corresponding polymer captions. Here we provide five examples of generated caption as reference:

"Poly(1-phenylethylene), commonly known as Polystyrene (PS), is a versatile thermoplastic with the molecular formula C8H8 and a formula weight of 104.15 g/mol. Synthesized via radical addition polymerization of styrene monomers at 70°C under a nitrogen atmosphere, PS boasts a distinctive aromatic structure with phenyl groups. This unique arrangement imparts rigidity, excellent electrical insulation, and optical clarity to the polymer. Its key properties include low cost, ease of processing, transparency, and good thermal stability, making PS indispensable in consumer goods like packaging, disposable cutlery, insulation materials, and as a substrate for optical applications. The polymer's affordability and versatility underscore its extensive use across industries, from packaging to electronics, where its ability to be molded into

complex shapes and its resistance to chemicals are highly valued."

"Poly(ethylene terephthalate) (PET), a high-performance polyester, is composed of repeating ethylene glycol and terephthalic acid units, with the molecular formula C10H8O4 and a formula weight of 192.17 g/mol. Synthesized through polycondensation of terephthalic acid and ethylene glycol, PET features ester linkages that contribute to its excellent mechanical strength, chemical resistance, and thermal stability. Its semi-crystalline structure enhances durability while maintaining transparency. Widely used in packaging, textiles, and engineering applications, PET is the material of choice for beverage bottles, food containers, synthetic fibers, and high-strength films due to its recyclability, lightweight nature, and resistance to moisture and wear."

"Poly(1-chloroethylene), or Polyvinyl Chloride (PVC), is a widely used polyvinyl polymer with the molecular formula C2H3Cl and a formula weight of 62.5 g/mol. It is *synthesized through the addition polymerization of vinyl chloride, PVC's structure includes chlorine atoms along the polymer backbone, contributing to its unique properties. Known for its high chemical resistance, excellent insulation capabilities, and durability, PVC is extensively used in construction for pipes and fittings, in electrical insulation, packaging, and as a key component in medical devices due to its biocompatibility and cost-effectiveness. Its ability to be plasticized enhances its flexibility, making it suitable for a wide range of applications from rigid structural elements to flexible tubing."*

"Poly(oxyethylene), commonly known as polyethylene oxide (PEO) or polyethylene glycol (PEG), is a versatile polymer within the polyoxides class, with the molecular formula C2H4O and a formula weight of 44.05 g/mol. Synthesized through the ring-opening polymerization of ethylene oxide, PEO features a linear chain with repeating ether units that impart exceptional solubility in water, flexibility, and biocompatibility. Its notable properties include high thermal stability, low toxicity, and the ability to

form hydrogen bonds, making it an excellent choice for applications requiring non-ionic surfactants, lubricants, and as a matrix for drug delivery systems. PEO's widespread use spans from pharmaceuticals, where it enhances drug solubility, to industrial applications like coatings, adhesives, and textiles, where its hydrophilic nature and low melting point are particularly beneficial."

"Poly(bisphenol A carbonate) (PC) is a high-performance polycarbonate with the molecular formula C16H14O3 and a formula weight of 254.28 g/mol. Typically synthesized through the reaction of 2,2-bis(4-hydroxyphenyl)propane (bisphenol A) with phosgene, PC features a robust backbone with carbonate linkages, providing exceptional impact resistance, thermal stability, and optical clarity. Known for its unique combination of toughness, transparency, and heat resistance, PC is extensively used in safety equipment, optical discs, electronic components, and automotive parts. Its ability to withstand extreme conditions while maintaining transparency and strength makes it invaluable in applications where both durability and visibility are crucial."

Pre-trained weights for encoders

Considering the limited amount of polymer-specific data, we utilized pre-trained weights from chemistry-related tasks for each encoder rather than training them from scratch. The weight files can be found in the following repositories:

SMILES: https://github.com/seyonechithrananda/bert-loves-chemistry

Graph: https://github.com/junxia97/Mole-BERT

Geometry: https://github.com/atomistic-machine-learning/schnetpack

Text Encoder: https://github.com/GT4SD/multitask_text_and_chemistry_t5

Hyperparameters Settings of the model

We summarize the hyperparameters as follows, with their corresponding symbols, as shown in Table S1.

Component	Hyperparameter	Symbol	Value
SMILES Encoder	Embedding Dimension	d_S	768
	Number of Attention Heads	n_S	12
	Number of Attention Layers	L_S	6
Graph Encoder	Node Embedding Dimension	d_G	300
	Edge Embedding Dimension	d_E	300
	Number of GNN Layers	L_G	5
	Node Attributes	X_G	Atom type, chirality
	Edge Attributes	E	Bond type, direction
Geometry Encoder	Atom Features	$X_\mathcal{G}$	Atomic number
	Atom Embedding Dimension	$d_\mathcal{G}$	128
	Number of Gaussian Functions	N_{Gauss}	50
	Number of Interaction Layers	$L_\mathcal{G}$	6
	Cutoff Radius	/	10 Å
	Number of Neighbors per Atom	N_{neigh}	32
Text Encoder	Embedding Dimension	d_C	768
	Number of Attention Layers	L_C	12
	Number of Attention Heads	n_C	12
Fingerprint Representation	Fingerprint Dimension	d_F	1024
	Fingerprint radius		1024
Projection Layer	Projection Layer Architecture	ρ	[*, 256] (‘*’ represents encoder output dimension)
Fusion Blocks	Number of Multi-Head Attention Heads	N_{head}	8
	Uni-poly Embedding Dimension	d	256

Pre-training	Temperature	τ	0.07
	Batch Size	/	32
	Epoch	/	10
	Learning Rate	/	1e-5
Fine-tuning	MLP Architecture	MLP	[256, 128, 64, 1]
	Batch Size	/	32
	Epoch	/	100
	Learning Rate	/	1e-4

Supplementary Table S1. Summary of the hyperparameters used in the model architecture and training process.

The training was conducted on an NVIDIA RTX A6000 GPU, with each batch taking approximately 1.34 seconds.

Training process

Pre-train loss

We use the InfoNCE as the pre-train loss. Formally, for a batch of N samples, given representations H_r and $H_{r'}$ for modalities r and r' $\left(H_r \in \mathbb{R}^d, H_{r'} \in \mathbb{R}^d\right)$, the similarity is defined using similarity:

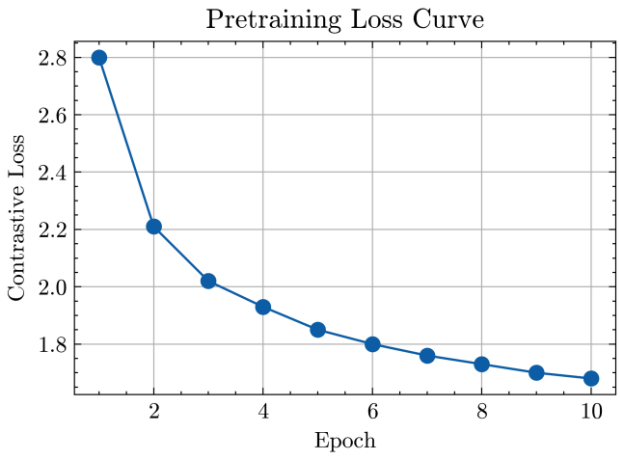
$$\mathrm{sim}\left(H_r^i, H_{r'}^j\right) = \frac{H_r^i \cdot H_{r'}^j}{\left|H_r^i\right|\left|H_{r'}^j\right|}.$$

The multimodal contrastive learning loss $\mathcal{L}$ is computed across all modality pairs, expressed as:

$$\mathcal{L} = \frac{1}{|\mathcal{P}|(|\mathcal{P}|-1)} \sum_{r \in \mathcal{P}} \sum_{r' \in \mathcal{P}, r \neq r'} \frac{1}{N} \sum_{i=1}^{N} -\log \frac{\exp\left(\frac{\mathrm{sim}\left(H_r^i, H_{r'}^i\right)}{\tau}\right)}{\sum_{j=1}^{N} \exp\left(\frac{\mathrm{sim}\left(H_r^i, H_{r'}^j\right)}{\tau}\right)}.$$

Here, τ is the temperature parameter that adjusts the dynamic range of similarity

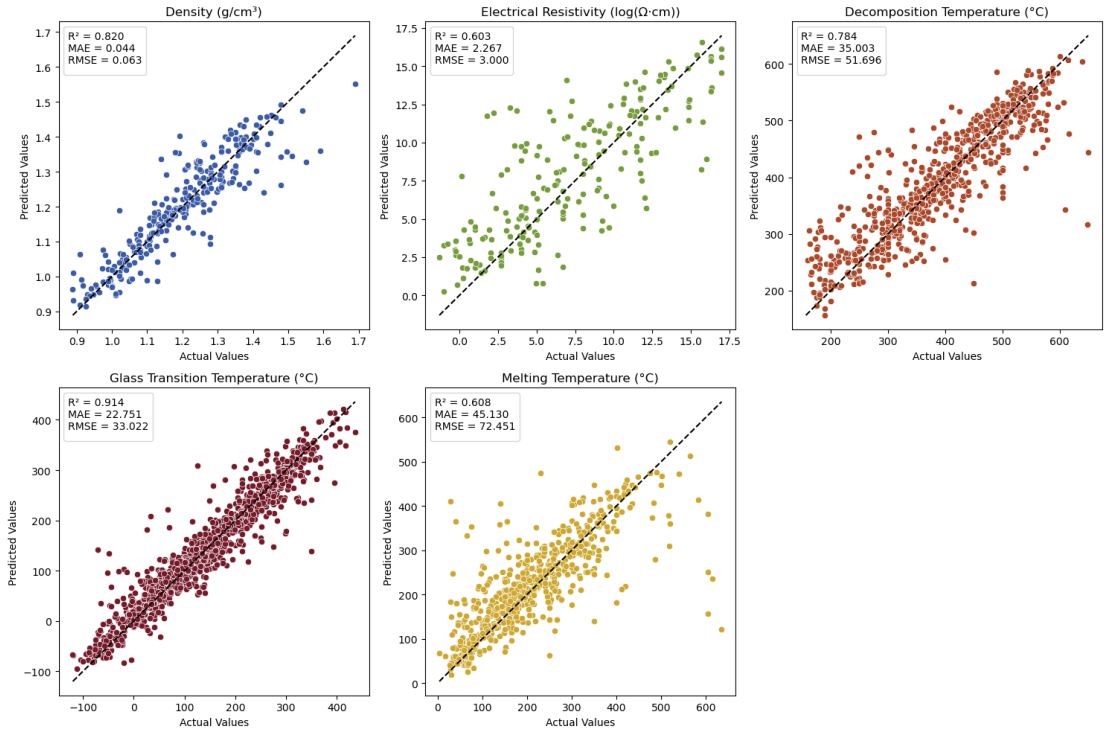
scores.



Supplementary Figure S1. Pretraining Loss Curve of Cross-Modal Contrastive Learning. The figure shows the contrastive loss (InfoNCE) over 10 epochs during the pretraining phase of the Uni-Poly model.

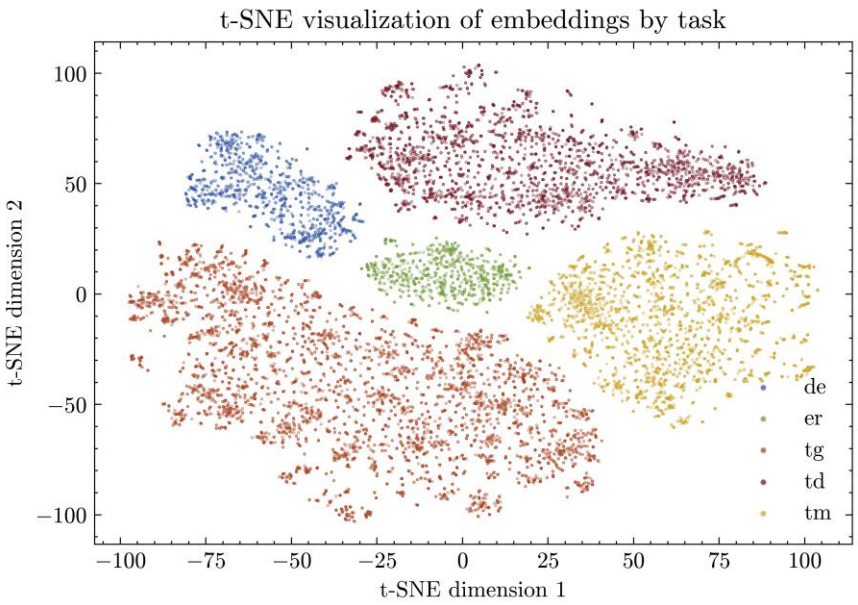
The loss curve, shown in Figure S1, illustrates a consistent decline over 10 epochs, decreasing from an initial value of 2.80 to a final value of 1.68. This steady reduction indicates effective learning of shared multimodal representations, providing a robust foundation for downstream tasks.

Predict results



Supplementary Figure S2. Predicted properties by the model versus experimental values. Scatter plots comparing predicted values to actual values for the test set across five properties: density, electrical resistivity, decomposition temperature, glass transition temperature, and melting temperature.

Figure S2 presents the predictive performance of the model on the test set for one of the five-fold cross-validation. Each scatter plot shows the relationship between the predicted and actual values for a specific property. The plots demonstrate varying levels of predictive accuracy, as indicated by R2, MAE, and RMSE metrics. The model shows strong predictive performance for glass transition temperature, whereas the performance for melting temperature and electrical resistivity is relatively lower.



Supplementary Figure S3. t-SNE Visualization of Uni-Poly Embeddings Clustered by Prediction Tasks. This figure presents a t-SNE scatter plot of the Uni-Poly embeddings, colored according to the prediction tasks corresponding to five polymer properties. Each point represents a polymer sample, and the clustering reflects the model's learned representation with respect to task-specific features.

Figure S3 provides a t-SNE visualization of embeddings obtained from five distinct models, each trained for predicting a specific polymer property. The embeddings for each property form well-defined clusters, indicating that each model effectively captures the unique feature space required for its corresponding task. This result highlights the specialization of individual models in learning property-specific representations and their ability to create task-differentiated embeddings. The distinct separations between clusters underscore the models' capability to focus on unique characteristics relevant to each polymer property.

Supplementary Reference

1. Giray, L. Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Ann Biomed Eng* 51, 2629–2633 (2023).

2. Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: Large-Scale Self-Supervised

Pretraining for Molecular Property Prediction. Preprint at http://arxiv.org/abs/2010.09885 (2020).

3. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How Powerful are Graph Neural Networks? in *International Conference on Learning Representations* (2018).

4. Schütt, K. *et al.* Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **30**, (2017).

5. Christofidellis, D. *et al.* Unifying molecular and textual representations via multi-task language modelling. in *International Conference on Machine Learning* 6140–6157 (PMLR, 2023).