Feedback-aligned Mixed LLMs for Machine Language-Molecule Translation

Dimitris Gkoumas[1*] **Maria Liakata**[1*]

[1]Queen Mary University of London
{d.gkoumas,m.liakata}@qmul.ac.uk

Abstract

The intersection of chemistry and Artificial Intelligence (AI) is an active area of research focused on accelerating scientific discovery. While using large language models (LLMs) with scientific modalities has shown potential, there are significant challenges to address, such as improving training efficiency and dealing with the out-of-distribution problem. Focussing on the task of automated language-molecule translation, we are the first to use state-of-the art (SOTA) human-centric optimisation algorithms in the cross-modal setting, successfully aligning cross-language-molecule modals. We empirically show that we can augment the capabilities of scientific LLMs without the need for extensive data or large models. We conduct experiments using only 10% of the available data to mitigate memorisation effects associated with training large models on extensive datasets. We achieve significant performance gains, surpassing the best benchmark model trained on extensive in-distribution data by a large margin and reach new SOTA levels. Additionally we are the first to propose employing non-linear fusion for mixing cross-modal LLMs which further boosts performance gains without increasing training costs or data needs. Finally, we introduce a fine-grained, domain-agnostic evaluation method to assess hallucination in LLMs and promote responsible use.

1 Introduction

Chemistry plays a crucial role in developing innovative scientific solutions that are scalable and cost-effective, whether it involves pioneering new drugs Ferguson and Gray [2018], advanced materials Kippelen and Brédas [2009], or improving chemical processes Zhong et al. [2023]. However, navigating the vast possibilities in chemistry but also biology and materials science necessitates the involvement of Artificial Intelligence (AI) technology to enable the acceleration of scientific discovery AI4Science and Quantum [2023].

Multi-modal models combining language with molecules have recently gained attention as a promising approach for studying and comprehending molecules, offering significant potential to tackle pressing global issues Zhang et al.. Existing work has applied successful paradigms from natural language processing (NLP) and multimodal representation learning to the chemistry domain. One common approach involves converting the inherent three-dimensional structures of molecules into SMILES, which provide a mapping to symbolic character-level representations. Subsequently, researchers have explored learning language-molecule representations in separate yet coordinated spaces Edwards et al. [2021, 2022], Liu et al. [2023a], in a joint space Liu et al. [2023b], or through hybrid approaches Luo et al. [2023], Christofidellis et al. [2023]. Despite significant advancements in the field, existing work does not effectively tackle the inherent challenges in training such models; they often rely on low quality or noisy synthetic data and typically require significantly more data compared to ordinary NLP tasks Edwards et al. [2021].

Submitted to NeurIPS 2024. Under review.

However training larger models and utilising extensive datasets does not always guarantee better performance. Ongoing research in NLP is focused on innovative training methods to enhance the performance of small-sized LLMs Liu et al. [2024]. For instance, ALMA Xu et al. [2023] fine-tuned LLaMA-2 Touvron et al. [2023] with extensive non-English monolingual data to enhance the model's multilingual abilities. Subsequently, supervised fine-tuning (SFT) with high-quality parallel data was performed to instruct the model in generating translations, surpassing all prior moderate-sized LLMs, including large models like GPT-3.5 Brown et al. [2020]. Another successful paradigm, Reinforcement Learning with Human Feedback (RLHF) Ouyang et al. [2022], aims to enhance the capabilities of LLMs across multiple NLP tasks, including machine translation. A recent trend is favouring closed-form losses that directly operate on offline preferences Rafailov et al. [2024], Xu et al. [2024], Ethayarajh et al. [2024], simplifying over RHLF without loss in performance.

Here, we address challenges of effectively training robust LLMs and dealing with the out-of-distribution problem when integrating scientific modalities, specifically focusing on machine language-molecule translation. Our approach is the first to experiment with human-centric optimisation objectives in the cross-modal representation learning setting to align cross language-molecule modals with preferred offline translations. To mitigate the memorisation effects inherent in training on large datasets, we conduct experiments using only 10% of the *L+M-24* benchmark dataset Edwards et al. [2024], achieving new state-of-the-art (SOTA) levels. Additionally, we propose a novel approach that fuses cross-modal LLMs to amplify performance gains without increasing training costs or data needs. The contributions of this work are as follows:

- We obtain robust LLMs for machine language-molecule translation by experimenting with recent shifts in RLHF within the cross-modal representation setting, namely, direct preference optimisation (DPO), contrastive preference optimisation (CPO), and Kahneman-Tversky optimisation (KTO) (§ 3.2). To the best of our knowledge, this is the first comprehensive study that investigates human-centric optimisations in LLMs in the multidisciplinary field of AI for living sciences. Our experiments show that both DPO and KTO cannot cope within agnostic cross-modal settings. In contrast, CPO, which is invariant to biases in these settings, achieves substantial performance gains by a large margin, reaching new SOTA levels compared to the best benchmark, Meditron, trained on the entire dataset (§ 4.4.1).
- We introduce an approach that mixes dominant-modality LLMs, based on non-linear fusion algorithms, namely TIES and SLERP (§ 3.3). This addresses inequalities between the informativeness of different modalities as well as inherent limitations in optimisations operating on preference data, all without increase in training costs. We empirically demonstrate that this approach facilitates RLHF optimisation algorithms, which, to varying degrees, depend on how well the pre-trained model is suited to specific cross-modal settings, thereby further boosting performance gains (§ 4.4.2).
- We introduce a fine-grained evaluation method with multi-aspect criteria for assessing the presence of hallucinations in LLMs. Our domain-agnostic method can help ensure the responsible use of chemical LLMs (§ 3.4). We critically discuss our new insights and compare them with established metrics in the literature. We conclude by highlighting the limitations of both our evaluation criteria and previous evaluation and discuss potential future directions for responsibly assessing generative outputs from chemical LLMs (see Section 4.4.3).

2 Foundations in RLHF for LLMs

Feedback-aligned LLMs traditionally undergo fine-tuning with RLHF, wherein human preferences serve as a reward signal in optimisation Stiennon et al. [2020], Ouyang et al. [2022]. To train a LLM with RLHF, a reinforcement learning optimisation algorithm such as PPO Schulman et al. [2017] is typically deployed on offline preference data, involving commonly three steps:

- **Roll-out:** Typically, a model π is trained for auto-regressive language generation on a large generic corpus. This training operates under the premise that the probability distribution of a sequence of words can be broken down into the product of conditional distributions for the next word Radford et al. [2019].
- **Evaluation:** A reference model π_{ref} is employed to optimise π for a downstream task. Typically, the π_{ref} model undergoes fine-tuning with an auto-regressive objective, using data pertinent to the downstream task. This often involves instruction tuning π_{ref} to regulate the generated outputs.
- **Optimisation:** The optimisation of π with respect to π_{ref} operates on a triple dataset $\mathcal{D} = \{x, y_w, y_l\}$, where x represents the input, and y_w and y_l denote preferred and dis-preferred outputs,

respectively, such that $y_w \succ y_l$ for x. In Bradley and Terry [1952], the probability of y_w being preferred over y_l in pairwise comparisons can be formulated as follows:

$$p^*(y_w \succ y_l|x) = \sigma(r^*(x, y_w) - r^*(x, y_l)) \tag{1}$$

Here, σ represents the logistic function, and r^* denotes the "true" reward function that underlies the preferences. As obtaining the true reward directly from a human would be prohibitively expensive, a reward model r_ϕ is trained to act as a surrogate. This is achieved by minimising the negative log-likelihood in human preference data;

$$\mathcal{L}(r_\phi) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}[\log \sigma(r_\phi(x, y_w) - r_\phi(x, y_l))] \tag{2}$$

Additionally, the Kullback-Leibler (KL) divergence between the outputs generated by π_{ref} and the parameterised π_θ models serves as an additional reward signal, ensuring that the generated responses closely align with the reference model. Consequently, an optimal model π_θ is one that maximises;

$$\mathbb{E}_{(x\in\mathcal{D},y\in\pi_\theta)}[r_\phi(x, y)] - \beta\mathcal{D}_{\text{KL}}(\pi_\theta(y|x)||\pi_{\text{ref}}(y|x)) \tag{3}$$

where β is the temperature parameter typically $\in [0.1, 0.5]$.

Human-aware Loss Functions (HALOs): RLHF can present challenges due to its inherent slowness and instability, especially in the case of highly varied outputs Zheng et al. [2024]. Recently, there has been a shift towards using closed-form losses in RLHF to align language models (LLMs) with human preferences. These losses are predominantly HALOs that model human biases, as discussed in Tversky and Kahneman [1992], aiming to maximise the margin between preferred and dis-preferred generated outputs (for the definition refer to Appx. B). This approach offers a mathematical equivalence with RLHF, while effectively addressing its inherent limitations.

3 Methodology

Existing approaches for chemical models heavily rely on expensive pre-training using large mono-modality datasets or per-task fine-tuning for each multi-modal task. Meanwhile, other efforts concentrate on multi-task learning, integrating vast multi-task datasets to enhance the capabilities of chemical models (for related work please refer to Appx. A). Here, we adopt a distinct approach by reducing reliance on extensive data and exploring optimisation algorithms and modality fusion techniques to enhance the capabilities of chemical LLMs.

3.1 Problem Definition: Machine Language-Molecule Translation (LMolT)

Let L denote the source language space and M denote the target molecule space. Consider a dataset $\mathcal{D}$, comprising pairs of source language sequences x represented in space L, and their corresponding target molecule sequences y in space M, represented as $\mathcal{D} = \{x^{(i)}, y^{(i)}\}_{i=1}^N$, where N is the total number of pairs. Here, we cast the problem as a cross-modal translation task, aiming to learn an optimal function $f : L \leftrightarrow M$ through a model π_θ parameterised by θ. We coordinate the language and molecule spaces through instructional modelling to regulate the translation process across both directions, i.e., $L \rightarrow M$ and $M \rightarrow L$ (please refer to Appx. C).

3.2 Feedback-Aligned Optimisations Algorithms

Here we investigate approximation policies for feedback alignment optimisation for the LMolT task.

Supervised Fine-Tuning (SFT): A model is optimised by minimising the disparity between the generated and intended target outputs. For a model parameterised by θ, denoted as π_θ, the loss function is defined as the negative log-likelihood of source-target pairs $(x, y) \sim \mathcal{D}$, and can be expressed as follows:

$$\mathcal{L}_{\text{NLL}} = -\mathbb{E}_{(x,y)\sim\mathcal{D}}[\log \pi_\theta(y|x)] \tag{4}$$

where the quality of (x, y) is crucial to the effectiveness of the translation model Zhao et al. [2022].

Direct Preference Optimisation (DPO): DPO Rafailov et al. [2024] provides a more human-centric optimisation objective aimed at aligning the translation model π with human intentions. Here $\mathcal{D}$ is an offline dataset of comparisons $\mathcal{D} = \{x^{(i)}, y_w^{(i)}, y_l^{(i)}\} i = 1^N$, where y_w represents the preferred translation (e.g. human gold standard), and y_l represents the dis-preferred translation (typically synthetic outputs obtained by a translation model). The loss function for DPO is constructed as a maximum likelihood objective for a parameterised policy π_θ;

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\Big[\log\sigma\underbrace{\Big(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\Big)}_{\text{Optimal reward } \forall(x,y)\in L\times M}\Big] \tag{5}$$

where π_{ref} is the fine-tuned translation model used as the reference, σ is the Sigmoid function, and β is a hyper-parameter (§ 2).

Contrastive Preference Optimisation (CPO): DPO requires high-quality data or, at the very least, appropriately fine-tuned models. In real-world scenarios, it may not be feasible to acquire high-quality data, especially for challenging domains such as chemistry and biology. CPO is a general approximation of Eq.5 that addresses the limitations of DPO by training a model to avoid generating outputs that are merely adequate, but not perfect, tested in machine translation Xu et al. [2024]. Specifically, in CPO, Eq. 5 is effectively approximated using a uniform reference model, which assumes equal likelihood for all possible generated outputs;

$$\mathcal{L}(\pi_\theta; U) = -\mathbb{E}_{(x,y_w,y_l)\sim D}\Big[\log\sigma\Big(\beta\log\pi_\theta(y_w|x) - \beta\log\pi_\theta(y_l|x)\Big)\Big] \tag{6}$$

Equation 6 implies that the loss is calculated based on how well generated outputs match this uniform distribution of possible outputs, rather than being biased towards any particular one. To maintain π_θ close to the preferred data distribution, a behaviour cloning (BC) Hejna et al. [2023] regulariser is injected;

$$\min_\theta \mathcal{L}(\pi_\theta, U) \quad \text{s.t.} \quad \mathbb{E}_{(x,y_w)\sim D}\Big[\mathbb{KL}(\pi_w(y_w|x)||\pi_\theta(y_w|x))\Big] < \epsilon, \tag{7}$$

Here, ϵ denotes a small positive constant, and $\mathbb{KL}$ signifies the Kullback-Leibler divergence. Finally, the regulariser is streamlined with an additional SFT term on the preferred data;

$$\mathcal{L}_{\text{CPO}} = \min_\theta \underbrace{\mathcal{L}(\pi_\theta, U)}_{\mathcal{L}_{\text{prefer}}} - \underbrace{\mathbb{E}_{(x,y_w)\sim D}[\log\pi_\theta(y_w|x)]}_{\mathcal{L}_{\text{NLL}}} \tag{8}$$

Kahneman-Tversky Optimisation (KTO): KTO Ethayarajh et al. [2024] combines the implicit reward under the RLHF objective with the concept of loss aversion from Kahneman-Tversky prospect theory Tversky and Kahneman [1992]. Unlike DPO and CPO, KTO expects a specific dataset format $\mathcal{D} = \{x^{(i)}, y^{(i)}, y_\lambda^{(i)}\}_{i=1}^N$, where x represents the source, y denotes the output, and y_λ indicates whether y is preferred or dis-preferred. Thus, KTO does not rely on preferences but rather on knowing the desirability of output y. Intuitively, KTO's loss is determined by a *value function* $v \to \mathbb{R}$ applied to a generated output z, with respect to a reference point z_{ref} and a *weighting function* w that assigns weights to the loss for preferred and dis-preferred generated outputs. More specifically, for training purposes, the value function is cast as a logistic function σ;

$$v_{\text{KTO}}(x, y; \beta) = \begin{cases} \sigma(r_{KTO}(x,y) - z_{ref}) & \text{if } y \sim y_\lambda = \text{preferred} \\ \sigma(z_{\text{ref}} - r_{\text{KTO}}(x,y)) & \text{if } y \sim y_\lambda = \text{dis-preferred} \end{cases} \tag{9}$$

Here, r_{KTO} is set to be the implicit reward under the RLHF objective in Eq 5;

$$r_{\text{KTO}}(x, y) = \beta\log\frac{\pi_\theta(y|x)}{\pi_{\text{ref}}(y|x)} \tag{10}$$

where β is a scale factor and z_{ref} is calculated in relation to all pairs (x', y') as an approximation of the optimal policy, rather than being computed individually for each pair (x, y);

$$z_{\text{ref}} = \mathbb{E}_{x'\sim\mathcal{D}}\left[\beta\mathbb{KL}(\pi_\theta(y'|x')||\pi_{\text{ref}}(y'|x'))\right] \tag{11}$$

Ultimately, the KTO loss boils down to the following function:

$$\mathcal{L}_{\text{KTO}}(\pi_\theta, \pi_{\text{ref}}) = \mathbb{E}_{x,y\sim\mathcal{D}}\left[w(y)(1 - v_{\text{KTO}}(x, y; \beta))\right] \tag{12}$$

where $w(y)$ represents two hyper-parameters that weigh the losses for preferred and dis-preferred completions, respectively, as follows:

$$w(y) = \begin{cases} \lambda_P & \text{if } y \sim y_\lambda = \text{preferred} \\ \lambda_D & \text{if } y \sim y_\lambda = \text{dis-preferred} \end{cases} \tag{13}$$

3.3 Mixed Cross-Modals Fusion

We propose a novel fusion approach that mixes cross-modal capabilities of LLMs excelling in unidirectional LMolT without increasing training costs or data needs. Our approach builds on early work in model merging, which linearly combines weights of pre-trained models into a unified, robust model Wortsman et al. [2022]. This addresses challenges related to disparities in the informativeness of different modalities and limitations of optimisation algorithms requiring LLMs trained on extensive data. We have employed the following non-linear fusion algorithms for mixing cross-modal LLMs :

- **Trim, Elect Sign & Merge (TIES)** Yadav et al. [2024]: This method addresses inference challenges caused by parameter interference from merging different models. It involves three steps: (1) resetting parameters that changed minimally during fine-tuning, (2) resolving sign conflicts, and (3) merging only the parameters aligned with the final agreed-upon sign.
- **Spherical Linear Interpolation (SLERP)** Shoemake [1985]: It is a geometric approach that blends two models while preserving their distinctive characteristics and curvature in high-dimensional spaces. The process consists of three steps: (1) normalising vectors to unit length, prioritising directions over magnitudes, (2) calculating the scale factor, which determines the interpolation factor and the angle between vectors across the fused models, and (3) weighting and summing the vectors to derive the interpolated representation in the fused model.

3.4 Evaluation Method for Assessing Hallucination in Chemical LLMs

Previous work has employed embedding representations Jaeger et al. [2018] to assess semantics in chemical-domain models Edwards et al. [2021], Christofidellis et al. [2023]. However, these approaches often require domain adaptation for out-of-distribution data Edwards et al. [2024] and might yield opaque and arbitrary outcomes Steck et al. [2024]. Here, in addition to performing evaluation according to standard generation metrics, we also introduce fine-grained multi-aspect evaluation criteria for assessing the presence of hallucinations[1] in generated outputs.

Specifically, for our language-to-molecule translation evaluation, we introduce the following metrics:

- Δ_{Len}**:** Measures the deviation in character length between generated and reference molecule pairs.
- **Chr-F** Popović [2015]: Utilises the F-score statistic for character n-gram matches between prediction-reference pairs. Here it assesses matches in translated molecules against their references by averaging the scores of unigram, bigram, and trigram matches. A higher Chr-F indicates better performance.
- **Win Rate**: Determines the percentage of successfully generated molecules within the test subset. Each sample in the test set is considered a win if it meets several strict criteria: a) Chr-F score > 0.3, b) absolute length deviation from the reference < 5, and c) a valid generated output adhering to standard molecular structure conventions[2].

For our molecule-to-language translation evaluation, we introduce similar metrics:

- Δ_{Len}**:** Calculates the token-length deviation in generated-reference pairs of texts.
- **Natural Language Inference (NLI)** Williams et al. [2018]: We utilise scores from an NLI model[3]. Specifically, we obtain the probabilities of predictions being entailed in references. Higher probabilities indicate better performance.
- **Win Rate**: Measures the percentage of successfully generated outputs within the test subset. A generated output, in regard to its corresponding reference, is considered a win if classified by the NLI model as entailment rather than neutral or contradiction.

[1] Hallucination in LLMs refers to a phenomenon where the generated outputs are inaccurate, nonsensical, or contradictory to the provided factual information.

[2] Threshold values are determined experimentally, while validity is automatically assessed using rdkit.

[3] We use *NLI-DeBERTa* based on He et al. [2021].

4 Experiments

4.1 Data

Experiments are conducted on the *L+M-24* benchmark dataset, which integrates molecular and linguistic modalities segmented into four categories with significant applications: biomedical; light and electricity; human interaction and organoleptics; and agriculture and industry Edwards et al. [2024]. This dataset is created by extracting molecules and their associated chemical properties from chemical databases. These properties are subsequently transformed into natural language using a template composition procedure with GPT-4.

The training and validation subsets comprise approximately 127k and 34k language-molecule pairs, respectively. We utilise 10% of these subsets for training and validation purposes. To implement optimisation algorithms using offline preference data, we construct tailored triple datasets (refer to §3.2) comprising both preferred and dis-preferred outputs. The preferred outputs are comprised of the golden references from *L+M-24*, while we generate the dis-preferred outputs synthetically using MolT5 Edwards et al. [2022]. For evaluation, we randomly selected 3k unseen pairs from a distinct dataset provided as supplementary resources for extensive training of chemical LLMs at the Language + Molecules @ ACL2024 workshop.[4].

4.2 Benchmark Models

We compare our results with established language-molecule models as reported in the literature:

- TxtChem-T5 Christofidellis et al. [2023]: A $T5_{XL}$ model trained on both linguistic and molecule modalities with a multi-task objective across various datasets, including the CheBI-20 dataset Edwards et al. [2022], similar to *L+M-24*.
- Chem-LLM Zhang et al. [2024]: An InternLM2-Base-7B model, trained on an extensive chemical domain knowledge dataset, with a direct preference optimisation objective Rafailov et al. [2024], achieving results comparable to GPT-4.
- Meditron Chen et al. [2023]: A Meditron-7B model fine-tuned on the entire *L+M-24* dataset for unidirectional language-molecule translation.
- SFT-Meditron: Our supervised Meditron-7B fine-tuned on a 10% subset of *L+M-24* for bi-directional machine language-molecule translation.

4.3 Experimental Setup

All experiments are conducted using Meditron Chen et al. [2023], a model previously noted for its performance on L+M+24. When investigating optimisation algorithms (§ 3.2), we initialise cross-modals from Meditron trained for unidirectional language-to-molecule translation[5]. In investigations of mixed modality fusion strategies (§ 3.3), we obtain cross-modals from separate Meditron models trained for unidirectional language-to-molecule and molecule-to-language translation, fusing them in a 19:1 ratio[6]. All models are trained with QLoRA Dettmers et al. [2024], with experimental settings detailed in Appx. F. For evaluation metrics please refer to Appx D

4.4 Results

4.4.1 Quantitative Analysis on Feedback-Aligned Optimisation Algorithms

Table 1 presents a summary of the molecule-to-language translation results. We observe a notable decrease in performance for benchmark models trained on extensive data with SFT when evaluated on unseen (out-of-distribution data). Among the baselines, Meditron exhibited the highest performance, likely due to its training on the entire *L+M-24* dataset used in our experiments. However, training with both DPO and KTO results in failure to translate molecules to language when initialised from agnostic cross-modals, such as those trained from language to molecule translation (§ 4.3). This

[4]Sampling is conducted from the test subset of the separate supplementary data provided for the Language + Molecules @ ACL2024 workshop; `https://github.com/language-plus-molecules/LPM-24-Dataset`

[5]Initialisation of cross-modals is based on the task reported as most challenging in Edwards et al. [2024].

[6]The experimental ratio of cross-modals aims to maintain information integrity across them.

suggests that these optimisation algorithms rely heavily on the model's suitability for a specific task. Our SFT demonstrated competitive performance when trained on a 10% subset of *L+M-24* with a bi-directional LMolT objective, but it underperformed compared to Meditron, trained on the entire dataset, and lacked efficiency (for efficiency details, refer to Appendix G). By contrast, CPO showed the ability to cope in the agnostic cross-modal setting, achieving up to a 20% performance increase compared to Meditron (see Table 1).

Method	Blue-2 ↑	Blue-4 ↑	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑	METEOR ↑
TxtChem-T5	0.08	0.09	0.19	0.06	0.17	0.16
Chem-LLM	0.03	0.00	0.11	0.02	0.09	0.14
Meditron	0.42	0.30	0.63	0.47	**0.49**	0.54
SFT	0.37	0.26	0.55	0.40	0.39	0.61
DPO	0.00	0.00	0.00	0.00	0.00	0.00
CPO	**0.62**	**0.45**	**0.68**	**0.50**	0.48	**0.62**
KTO	0.00	0.00	0.00	0.00	0.00	0.00
$\Delta_{CPO\text{vs}MED}$	+20%	+19%	+5%	+3%	-1%	+8%

Table 1: Molecule-to-Language Translation Results on 3k Unseen Pairs. Arrows next to metrics indicate the higher value the better performance. Best results are highlighted in bold. $\Delta_{CPO\text{vs}MED}$ represents the performance gain of our best model compared to Meditron trained on the entire data.

Performance gains in language-to-molecule translation were notably higher when cross-modals were known (see Table 2). Both DPO and CPO demonstrated similar performance, achieving an increased accuracy of up to 42% compared to Meditron trained on the entire dataset. Conversely, KTO exhibited low performance even in a known cross-modal setting. We suspect that KTO is sensitive to overfitting (refer to Appendix G).

Method	BLEU ↑	Levenshtein ↓	MACCS FTS ↑	RDK FTS ↑	Morgan FTS ↑	FCD ↓	Validity ↑
TxtChem-T5	0.18	133.29	0.21	0.10	0.03	37.67	0.58
Chem-LLM	0.04	732.74	0.00	0.00	0.00	59.44	0.19
Meditron	0.43	66.16	0.35	0.29	0.19	13.64	0.57
SFT	0.30	186.99	0.70	0.62	0.41	11.14	0.98
DPO	**0.72**	**42.40**	**0.77**	0.69	**0.49**	10.47	0.99
CPO	0.71	42.65	**0.77**	**0.70**	0.48	**4.19**	**1.00**
KTO	0.23	294.63	0.03	0.03	0.02	32.64	0.06
$\Delta_{CPO\text{vs}MED}$	+29%	-23.76%	+42%	+41%	+30%	-9.45%	+41%

Table 2: Language-to-Molecule Translation Results on 3k Unseen Pairs. Arrows next to metrics indicate whether higher or lower values denote better performance. Best results are highlighted in bold. $\Delta_{CPO\text{vs}MED}$ represents the performance gain of our best models compared to Meditron.

In summary, our experiments revealed that CPO operates independently of cross-modal considerations, whereas DPO cannot cope with an agnostic cross-modal setting. Contrary, KTO displayed signs of overfitting even in a known cross-modal setting. In the following section, we explore fusion approaches aimed at enhancing the capabilities of these algorithms within the cross-modal context.

4.4.2 Quantitative Analysis on Mixed Cross-Modals Fusion

Her we focus on the most effective optimisation algorithms, DPO and CPO (§ 4.4.1), and investigate TIES and SLERP (§ 3.3) fusion algorithms in the cross-modal context. We thus aim to address disparities between language and molecule modalities and enhance optimisation algorithm performance. Tables 3 and 4 provide an overview of the mixed cross-modal fusion results for LMolT, maintaining consistent training data. Combining DPO with mixed cross-modals via TIES shows promising enhancements in molecule-to-language translation (see $\Delta_{DPO\text{vs}TIES+DPO}$ in Table 3), yet significant performance loss in language-to-molecule translation (see $\Delta_{DPO\text{vs}TIES+DPO}$ in Table 4). Conversely, fusing CPO with mixed cross-modals via SLERP notably improves molecule-to-language capabilities (see $\Delta_{CPO\text{vs}SLERP+CPO}$ in Table 3) while minimally impacting language-to-molecule translation performance (see $\Delta_{CPO\text{vs}SLERP+CPO}$ in Table 4), showcasing overall gains compared

to Meditron trained on the entire dataset. Overall, we consistently observe a trade-off in performance when fusing cross-modals, prompting future exploration of nuanced algorithms for modelling cross language-molecule dynamics.

Fusion	Method	Blue-2 ↑	Blue-4 ↑	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑	METEOR ↑
TIES	**DPO**	0.74	0.53	0.74	0.54	0.51	0.70
	CPO	0.74	0.54	0.76	0.57	0.53	0.72
SLERP	DPO	0.00	0.00	0.02	0.01	0.00	0.00
	CPO	0.73	0.53	0.76	0.56	0.53	0.71
$\Delta_{DPO vs TIES+DPO}$		+74%	+53%	+74%	+54%	+51%	+70%
$\Delta_{CPO vs SLERP+CPO}$		+11%	+8%	+8%	+6%	+5%	+9%
$\Delta_{MED vs SLERP+CPO}$		+31%	+28%	+13%	+9%	+4%	+17%

Table 3: Mixed Cross-Modals Fusion Results in Molecule-to-Language Translation. Best combined approaches are highlighted in bold. $\Delta_{DPO vs TIES+DPO}$, $\Delta_{CPO vs SLERP+CPO}$, and $\Delta_{MED vs SLERP+CPO}$ measure performance gains of the best-combined approaches from the singular cross-modal setting of *DPO*, *CPO*, and the benchmark *Meditron*, as reported in Table 1.

Fusion	Method	BLEU ↑	Levenshtein ↓	MACCS FTS ↑	RDK FTS ↑	Morgan FTS ↑	FCD ↓	Validity ↑
TIES	**DPO**	0.32	93.18	0.31	0.22	0.19	19.80	0.42
	CPO	0.68	46.91	0.72	0.65	0.45	24.50	0.94
SLERP	DPO	0.72	43.85	0.77	0.70	0.51	10.35	0.98
	CPO	0.71	44.01	0.73	0.66	0.45	11.22	0.95
$\Delta_{DPO vs TIES+DPO}$		-40%	+51%	-46%	-47%	-30%	+7.33%	+58%
$\Delta_{CPO vs SLERP+CPO}$		0%	+1.36%	-4%	-4%	-3%	+5%	-4%
$\Delta_{MED vs SLERP+CPO}$		+29%	-22.40%	+38%	+37%	+27%	-4.45%	+37%

Table 4: Mixed Cross-Modals Fusion Results in Language-to-Molecule Translation. Best combined approaches are highlighted in bold. $\Delta_{DPO vs TIES+DPO}$, $\Delta_{CPO vs SLERP+CPO}$, and $\Delta_{MED vs SLERP+CPO}$ measure performance gains of the best-combined approaches from the singular cross-modal setting of *DPO*, *CPO*, and the benchmark *Meditron*, as reported in Table 1.

4.4.3 Evaluation Results on Assessing Hallucination in Chemical LLMs

Hallucinations can negatively impact the usability of LMoT models. Figure 1 illustrates the distribution of metrics introduced in § 3.4 for our best-performing models, CPO and SLERP+CPO, in different cross-modal settings compared to Meditron, which was trained on the entire dataset. In the cross-language-to-molecule setting, unlike Meditron, which tends to generate molecules shorter by an average of 27 characters, our models produced molecules with lengths similar to the actual ones. When evaluating Char-F, we observed a substantial rightward shift in distributions (Fig. 1 (A)), indicating that our models are significantly better at capturing uni-, bi-, and tri-modal character-level dynamics in molecules. In the molecule-to-language setting, our models generated language descriptions of similar length to the actual ones, whereas Meditron consistently produced shorter predictions. When measuring entailment probabilities, we observed a decrease in distributions on the left side and an increase on the right side for our models (Fig. 1 (B)), indicating that our models understand underlying concepts rather than merely memorising language patterns. Figure 2 (B) illustrates a case study of this phenomenon. Additional examples can be found in Appendix E.

We proceeded by selecting models that competed with Meditron in at least one cross-modal setting and conducted a cross win rate comparison among them, as illustrated in Figure 2. For language-to-molecule translation, we observed a decline in molecule generation performance for Meditron and SFT when evaluated across multiple criteria, as outlined in § 3.4. Conversely, CPO and SLERP+CPO demonstrated success in the multi-aspect evaluation, although previous SLERP+CPO showed a slight performance decrease when evaluated on single-aspect benchmark metrics (see $\Delta_{CPO vs SLERP+CPO}$ in Table 4). For molecule-to-language translation, the NLI win rates (§ 3.4) of SFT and TIES+DPO contradicted with the performance as reported in Tables 1 and 3, while CPO and SLERP+CPO exhibited consistency with the evaluation metrics in those tables. Although NLI evaluation generally surpasses statistical methods when assessing generated outputs against expected references, it still faces limitations, including accuracy issues with lengthy texts and dependence on the quality of

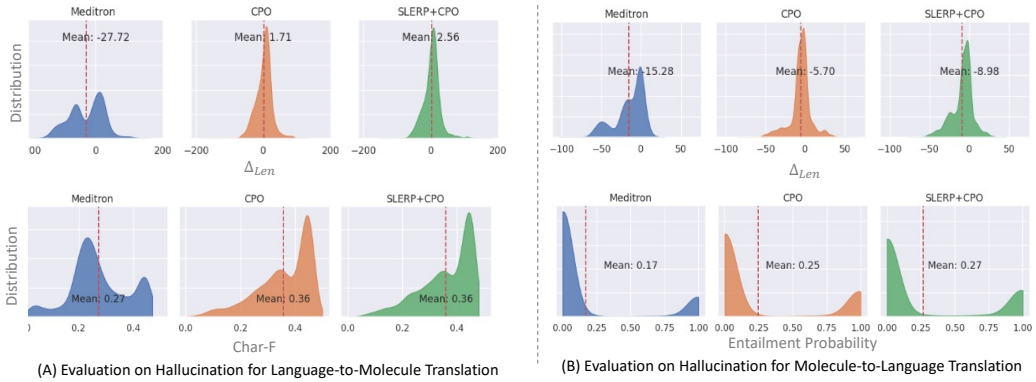

Figure 1: Distributions of hallucination measurements (§ 3.4) across 3000 pairs for our best language-molecule translation models trained on 10% of data against the benchmark Meditron trained on the entire one. (A) Hallucination in language-to-molecule translation: A Δ_{Len} closer to zero and higher Char-F scores indicate superior performance. (B) Hallucination in molecule-to-language translation: A Δ_{Len} closer to zero and higher entailment probabilities indicate superior performance.

training dataMcIntosh et al. [2024]. In the future, we aim to explore more nuanced methods that utilise appropriate LLMs for evaluating generated language Min et al. [2023].

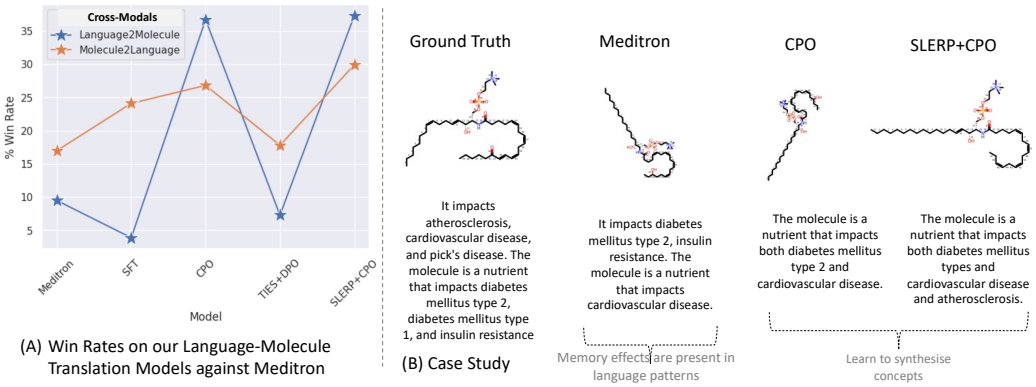

Figure 2: (A) Win rate of top-performing models in at least one cross-modal setting against Meditron. (B) A case study demonstrating that our best models learn to synthesise concepts rather than memorising patterns. More samples available in Appx. E

5 Conclusion

We focuse on the machine language-molecule translation, aiming to empirically tackle the challenges of effectively training robust large language models (LLMs) and addressing the out-of-distribution problem when integrating with scientific modalities. Our approach marks the first endeavour to experiment with human-centric optimisation algorithms in the context of cross-modal representation learning, aligning cross-language-molecule modalities with preferred offline translations. To combat memorisation effects inherent in training large models on extensive datasets, we conduct experiments utilising only 10% of the available data. This strategy yields significant performance enhancements, surpassing the best benchmark model trained on extensive data by a wide margin, thus achieving new state-of-the-art levels. Additionally, we are the first to propose employing non-linear fusion for integrating cross-modal LLMs, addressing challenges related to disparities in the informativeness of different modalities and limitations of optimisation algorithms requiring LLMs trained on extensive data. This approach further enhances performance gains without increasing training costs or data requirements. Finally, we introduce a freely-tied domain evaluation method with multi-aspect criteria for assessing the presence of hallucinations in chemical LLMs. In the future, we aim to investigate

more nuanced algorithms for capturing language-molecule dynamics and evaluation methods to promote their responsible use.

References

Microsoft Research AI4Science and Microsoft Azure Quantum. The impact of large language models on scientific discovery: a preliminary study using gpt-4. *arXiv preprint arXiv:2311.07361*, 2023.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Zeming Chen, Alejandro Hernández Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, et al. Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*, 2023.

Dimitrios Christofidellis, Giorgio Giannone, Jannis Born, Ole Winther, Teodoro Laino, and Matteo Manica. Unifying molecular and textual representations via multi-task language modelling. In *International Conference on Machine Learning*, pages 6140–6157. PMLR, 2023.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36, 2024.

Carl Edwards, ChengXiang Zhai, and Heng Ji. Text2mol: Cross-modal molecule retrieval with natural language queries. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 595–607, 2021.

Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. Translation between molecules and natural language. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 375–413, 2022.

Carl Edwards, Qingyun Wang, Lawrence Zhao, and Heng Ji. L+ m-24: Building a dataset for language+ molecules@ acl 2024. 2024.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. Kto: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.

Fleur M Ferguson and Nathanael S Gray. Kinase inhibitors: the road ahead. *Nature reviews Drug discovery*, 17(5):353–377, 2018.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*, 2021.

Joey Hejna, Rafael Rafailov, Harshit Sikchi, Chelsea Finn, Scott Niekum, W Bradley Knox, and Dorsa Sadigh. Contrastive prefence learning: Learning from human feedback without rl. *arXiv preprint arXiv:2310.13639*, 2023.

Sabrina Jaeger, Simone Fulle, and Samo Turk. Mol2vec: unsupervised machine learning approach with chemical intuition. *Journal of chemical information and modeling*, 58(1):27–35, 2018.

Bernard Kippelen and Jean-Luc Brédas. Organic photovoltaics. *Energy & Environmental Science*, 2 (3):251–261, 2009.

Alisa Liu, Xiaochuang Han, Yizhong Wang, Yulia Tsvetkov, Yejin Choi, and Noah A Smith. Tuning language models by proxy. *arXiv preprint arXiv:2401.08565*, 2024.

Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Chaowei Xiao, and Animashree Anandkumar. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12):1447–1457, 2023a.

Zequn Liu, Wei Zhang, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Ming Zhang, and Tie-Yan Liu. Molxpt: Wrapping molecules with text for generative pre-training. *arXiv preprint arXiv:2305.10688*, 2023b.

Jieyu Lu and Yingkai Zhang. Unified deep learning model for multitask reaction predictions with explanation. *Journal of chemical information and modeling*, 62(6):1376–1387, 2022.

Yizhen Luo, Kai Yang, Massimo Hong, Xingyi Liu, and Zaiqing Nie. Molfm: A multimodal molecular foundation model. *arXiv preprint arXiv:2307.09484*, 2023.

Timothy R McIntosh, Teo Susnjak, Tong Liu, Paul Watters, and Malka N Halgamuge. Inadequacies of large language model benchmarks in the era of generative artificial intelligence. *arXiv preprint arXiv:2402.09880*, 2024.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12076–12100, 2023.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Maja Popović. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, pages 392–395, 2015.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. 2019.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A Hunter, Costas Bekas, and Alpha A Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science*, 5(9):1572–1583, 2019.

Molecular Sets. A benchmarking platform for molecular generation models., 2022.

Ken Shoemake. Animating rotation with quaternion curves. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, pages 245–254, 1985.

Harald Steck, Chaitanya Ekanadham, and Nathan Kallus. Is cosine-similarity of embeddings really about similarity? *arXiv preprint arXiv:2403.05440*, 2024.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021, 2020.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5:297–323, 1992.

Alain C Vaucher, Federico Zipoli, Joppe Geluykens, Vishnu H Nair, Philippe Schwaller, and Teodoro Laino. Automated extraction of chemical synthesis actions from experimental procedures. *Nature communications*, 11(1):3601, 2020.

Adina Williams, Nikita Nangia, and Samuel Bowman. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122, 2018.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pages 23965–23998. PMLR, 2022.

Haoran Xu, Young Jin Kim, Amr Sharaf, and Hany Hassan Awadalla. A paradigm shift in machine translation: Boosting translation performance of large language models. *arXiv preprint arXiv:2309.11674*, 2023.

Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation. *arXiv preprint arXiv:2401.08417*, 2024.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in Neural Information Processing Systems*, 36, 2024.

Di Zhang, Wei Liu, Qian Tan, Jingdan Chen, Hang Yan, Yuliang Yan, Jiatong Li, Weiran Huang, Xiangyu Yue, Dongzhan Zhou, et al. Chemllm: A chemical large language model. *arXiv preprint arXiv:2402.06852*, 2024.

X Zhang, L Wang, J Helwig, Y Luo, C Fu, Y Xie, M Liu, Y Lin, Z Xu, K Yan, et al. Artificial intelligence for science in quantum, atomistic, and continuum systems. arxiv 2023. *arXiv preprint arXiv:2307.08423*.

Yao Zhao, Mikhail Khalman, Rishabh Joshi, Shashi Narayan, Mohammad Saleh, and Peter J Liu. Calibrating sequence likelihood improves conditional language generation. In *The Eleventh International Conference on Learning Representations*, 2022.

Chen Zheng, Ke Sun, Hang Wu, Chenguang Xi, and Xun Zhou. Balancing enhancement, harmlessness, and general capabilities: Enhancing conversational llms with direct rlhf. *arXiv preprint arXiv:2403.02513*, 2024.

Ming Zhong, Siru Ouyang, Yizhu Jiao, Priyanka Kargupta, Leo Luo, Yanzhen Shen, Bobby Zhou, Xianrui Zhong, Xuan Liu, Hongxiang Li, et al. Reaction miner: An integrated system for chemical reaction extraction from textual data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 389–402, 2023.

A Related Work

Specialised models tailored for chemistry have emerged, including Molecular Transformers Schwaller et al. [2019] and the RXN family Vaucher et al. [2020], which tackle tasks such as forward reaction prediction and molecular retrosynthesis. Yet, these models necessitate distinct architectures for individual tasks, resulting in heightened computational demands and a prerequisite for domain-specific proficiency in each task.

T5Chem Lu and Zhang [2022] introduces a cohesive multi-tasking framework for chemistry, leveraging a singular model for tasks spanning reaction prediction, regression, and classification. Nonetheless, T5Chem's reliance on task-specific heads confines its utility to the chemical realm, thus constraining its adaptability across diverse sub-domains.

MolT5 Edwards et al. [2022] confronts the intricate challenge of cross-domain generation by bridging natural language and chemistry, engaging in tasks like text-conditional de novo molecule generation and molecule captioning. Nonetheless, MolT5's effectiveness is curtailed by its dependence on resource-intensive pre-training on extensive mono-modality datasets, alongside per-task fine-tuning

for each multi-modal task. Consequently, this approach restricts MolT5's potential to fully exploit the multi-tasking capabilities of T5 and to facilitate seamless information sharing between tasks.

A multi-task T5 model underwent training with a multi-task objective tailored to handle tasks across various domains, including chemistry-based tasks (mol2mol), textual-based tasks (text2text), and cross-domain tasks (mol2text and text2mol) Christofidellis et al. [2023]. However, the majority of these approaches focus on optimising specific tasks, with all of them heavily relying on supervised fine-tuning using extensive datasets.

B Human-aware Loss Functions (HALOs)

Definition 1 (HALOs) *Let $x \in X$ and $y \in Y$ denote an input and output respectively. An $f : (x, y) \to \mathbb{R}$ is considered a human-aware loss function if it satisfies*

$$f(x, y; \theta) = t\Big(v_f(r_\theta(x, y) - \mathbb{E}_{x' \sim Q', y' \sim Q'}[r_\theta(x', y')])\Big) \tag{14}$$

with a parameterised reward function r_θ such that $\forall (x_1, y_1), (x_2, y_2) \in X \times Y$, $r_\theta(x_1, y_1) > r_\theta(x_2, y_2) \Leftrightarrow (x_1, y_1) \succ_{r_\theta} (x_2, y_2)$, reference point distributions $Q_x(X')$ and $Q_y(Y'|X')$, a value function $v_f : \mathbb{R} \to \mathbb{R}$ that is monotonic non-decreasing and concave in $(0, \infty)$, and a negative affine function t.

C Language-molecule Translation Instructions

Below is an instruction that describes a task, paired with an input that provides further context.
Write a response that appropriately completes the request.

Instruction: You are a researcher. You can come up captions based on your existing knowledge.
Captions are given against the following input. You should be as detailed as possible.

Input: Molecule: {source molecule}
In that molecule, could you formulate a caption about?

Response:{target caption}

Instruction for molecule to language translation, i.e., $M \to L$

Below is an instruction that describes a task, paired with an input that provides further context.
Write a response that appropriately completes the request.

Instruction: You are a researcher. You can come up molecule smile strings based on your existing knowledge.
Molecule smile strings are given against the following input. You should be as detailed as possible.

Input: Caption: {source caption}
In that caption, could you generate a molecule smile string?

Response: {target molecule}

Instruction for language to molecule translation, i.e., $L \to M$

D Evaluation Metrics

For performance evaluation, we employ established metrics from the literature Sets [2022], Edwards et al. [2022]. In molecule-to-language translation, we assess using BLEU-2, BLEU-4, ROUGE-1, ROUGE-2, ROUGE-L, and METEOR metrics. For language-to-molecule translation, evaluation metrics include BLEU, Levenshtein distance, fingerprint metrics (MACCS, RDK, and Morgan), Fréchet ChemNet Distance (FCD), and molecule validity metrics. Annotations in the result tables indicate whether higher or lower values indicate superior performance.

E Examples of Language-Molecule Translation

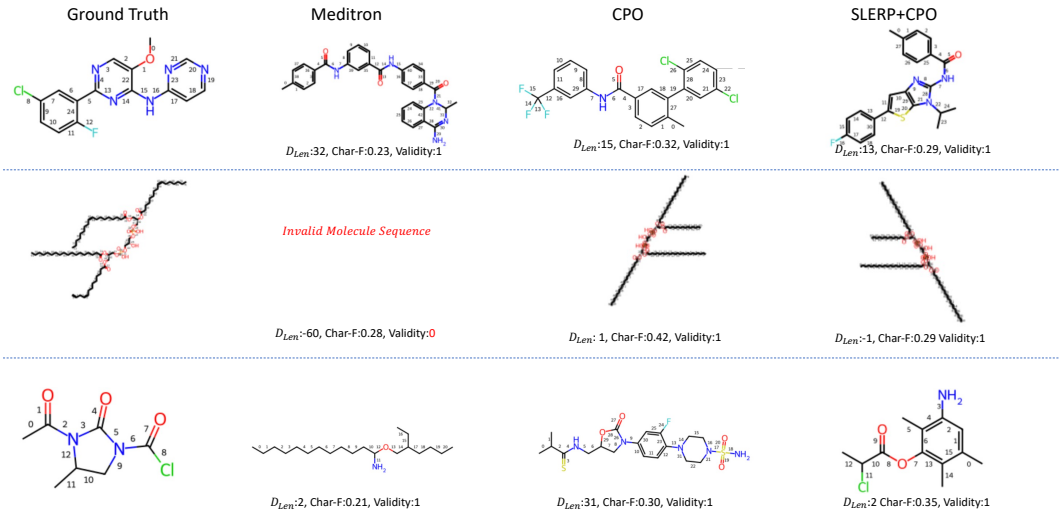

Figure 3: Examples of molecules generated by our top-performing models compared to the benchmark model, Meditron.

F Experimental Settings

Figure 4: Examples of generated language by our top-performing models compared to the benchmark model, Meditron.

G Training Efficiency

```
args = TrainingArguments(
    output_dir=save_path,
    overwrite_output_dir=True,
    load_best_model_at_end=True,
    num_train_epochs=3,
    per_device_train_batch_size=1
    per_device_eval_batch_size=1
    gradient_accumulation_steps=64
    gradient_checkpointing=False
    optim="adamw_torch_fused",
    learning_rate=5e-5,
    max_grad_norm=0.3,
    warmup_ratio=0.1,
    lr_scheduler_type="cosine",

)

(
    load_in_4bit=True,
    bnb_4bit_use_double_quant=True,
    bnb_4bit_quant_type=nf64,
      bnb_4bit_compute_dtype=torch.bfloat16
)

(
    lora_alpha=16,
    r = 64,
    lora_dropout=0.1,
    task_type="CAUSAL_LM",
    bias=False,
    target_modules= "all-linear"
)
```

Figure 5: Training configuration

Training efficiency across experimental methods