electronics

MDPI

Article

Romanian Style Chinese Modern Poetry Generation with Pre-Trained Model and Direct Preference Optimization

Li Zuo [1,2], Dengke Zhang [1], Yuhai Zhao [1] and Guoren Wang [3,*]

[1] School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China; zuoli@mail.neu.edu.cn (L.Z.); zhangdk@mail.neu.edu.cn (D.Z.); zhaoyuhai@mail.neu.edu.cn (Y.Z.)
[2] Foreign Studies College, Northeastern University, Shenyang 110819, China
[3] School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
* Correspondence: wanggrbit@126.com

Abstract: The poetry of distant country with different culture and language is always distinctive and fascinating. Chinese and Romanian belong to Sinitic languages of the Sino-Tibetan language family and Romance languages of the Indo-European language family, which have relatively different syntax and general imagery of literature. Therefore, in this study, we make an attempt that was rarely involved in previous poetry generation research, using modern Chinese as the carrier, and generating modern poetry with Romanian style based on pre-trained model and direct preference optimization. Using a 5-point grading system, human evaluators awarded scores ranging from 3.21 to 3.83 across seven evaluation perspectives for the generated poems, achieving 76.2% to 91.6% of the comparable scores for the Chinese translations of authentic Romanian poems. The coincidence of the 30th to the 50th most frequently occurring poetic images in both generated poems and Romanian poems can reach 58.0–63.3%. Human evaluation and comparative statistical results on poetic imagery show that direct preference optimization is of great help in improving the degree of stylization, and the model can successfully create Chinese modern poems with Romanian style.

Keywords: Romanian poetry; Chinese poetry; large language model; language style; natural language processing

check for updates

Academic Editors: Ioannis Hatzilygeroudis and Arkaitz Zubiaga

Received: 24 October 2024
Revised: 13 December 2024
Accepted: 9 January 2025
Published: 13 January 2025

Citation: Zuo, L.; Zhang, D.; Zhao, Y.; Wang, G. Romanian Style Chinese Modern Poetry Generation with Pre-Trained Model and Direct Preference Optimization. *Electronics* **2025**, *14*, 294. https://doi.org/10.3390/electronics14020294

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Language is the symbol and foundation of human civilization. From the legend of the Tower of Babel to the Rosetta Stone, from Sumerian cuneiform to inscriptions on bones or tortoise shells of the Shang Dynasty, language carries culture and history, leaving an indelible mark on the history of human civilization. In the process of human language interaction, whether spoken or written, "understanding" and "expression" are two indispensable parts. When we use machines to simulate human language interaction, "understanding" and "expression" correspond to "natural language understanding" and "natural language generation".

Natural Language Generation (NLG) is a significant research area within Natural Language Processing (NLP), and it also plays a crucial role in Artificial Intelligence (AI) research. NLG tasks can be divided into non-open end language generation tasks and open end language generation tasks. For example, machine translation and text summarization are non-open-end generation tasks, these tasks have relatively fixed answers and evaluation standards. Poetry generation, dialogue generation and story generation are

open-end generation tasks, they have high requirements for the innovativeness of the generated results.

As a very creative and interesting work in the field of NLP, poetry generation research has a long history. Like machine translation, poetry generation has similarly transitioned from rule-based and template-driven creation methods [1,2], to approaches grounded in statistical machine learning [3,4], and subsequently to systems powered by neural networks [5,6]. Recently, Large Language Models (LLMs) [7–9] have come to the forefront, and poetry generation based on pre-trained LLMs [10,11] has provided greater opportunities for improving the quality of generated poems.

Poetry can be generated by theme, format, or poet's style... and generating poetry with a specific language style is a new direction we want to try. On the one hand, poetry that uses one language to generate another language style blends the two languages in a clever way, greatly broadening the boundaries of the language and bringing novel feelings. On the other hand, the generation of stylized poetry is also very beneficial to the protection and continuation of minority culture. In the information age, with the help of the convenient Internet, information is generated and transmitted explosively, which makes the large languages that already have an overwhelming advantage continue to spread and iterate in the rapidly updating language environment, further squeezing the living space of small languages. Since small languages have relatively fewer users and a smaller total amount and increment of literature, the dissemination and development of small languages in the information society is more difficult than that of big languages. The exploration of the generation of poetry in small language styles will help improve the viability of small languages and protect cultural diversity.

In this paper, we comprehensively explore the use of neural networks for poetry generation by leveraging pre-trained LLMs to create modern Chinese poetry, fine-tune them to embody Romanian style, and employ Direct Preference Optimization [12] for further training to generate modern Chinese poetry characterized by a unique Romanian style.

In Section 2, we review existing studies on poetry generation and reinforcement learning (RL), outlining the reasons for our methodological choices. Section 3 details our method for generating Chinese modern poetry infused with Romanian stylistic elements. The experimental datasets and data collection process are described in Section 4. We present the experimental results and evaluate their effectiveness in Section 5. By summarizing our findings, Section 6 discusses future research directions. This paper aims to offer a thorough understanding of our work, guiding researchers in the poetry generation domain.

2. Related Work

2.1. Poetry Generation

Researchers from various countries have conducted a lot of research on poetry generation in various languages including English [13,14], Chinese [15–19], Japanese [20,21], French [10], Arabic [11,22], Spanish [1,23], Portuguese [24], Finnish [25,26], Bengali [27], Urdu and Hindi [28]. In addition to generating poems on specific topics [14], some researchers have also studied poetry generation methods for specific poets' styles [13]. In addition, researchers have also explored lyrics generation [29–31], which has certain similarities with poetry in language expression.

Chinese poetry is divided into three types: the form of pre-Tang poetry (古体诗), the regulated verse (近体诗) and the modern poetry (现代诗). Ancient Chinese is employed in the composition of pre-Tang poetry and regulated verse, imposing specific rules regarding rhyme schemes and the exact character count per line; while modern poetry is very free and has no requirements for meter and rhyme. As for the generation of Chinese poetry, according to whether the language used in the poetry is ancient Chinese or modern Chinese,

and whether the format followed is ancient poetry or modern poetry, it can be divided into "generation of ancient Chinese poetry" and "generation of modern Chinese poetry". The majority of prior research endeavors pertaining to the generation of Chinese poetry have centered around ancient poetry creation [15–18], with comparatively limited investigations into modern poetry generation [19]. The content of previous Chinese poetry generation mainly involves poetic images and emotions related to the Chinese context. Currently, there is scant research on the generation of Chinese poetry in alternative language styles. Chinese and Romanian belong to the Chinese group of the Sino-Tibetan language family and the Romance group of the Indo-European language family, which have relatively different syntax and general imagery of literature. Romanian poetry has a profound historical background and a very distinctive literary style. Therefore, in this study, we make an attempt that was rarely involved in previous poetry generation research, using Chinese as the medium to create poetry with a Romanian style.

In this study, when selecting the types of poems produced by the model, on the one hand we considered that both the form of pre-Tang poetry and the regulated verse are written in ancient Chinese, which makes it difficult to express contemporary things and Romanian allusions; on the other hand, considering that native Chinese speakers' impression of "Romanian style" mainly comes from literary works, dramas, and films translated in modern Chinese, we decided to generate Romanian-style Chinese poetry in the form of modern poetry.

2.2. Direct Preference Optimization

Direct Preference Optimization (DPO) is based on extensive research in reinforcement learning with human feedback (RLHF) and the modeling of human preferences. Traditional RLHF methods, such as those introduced by Ziegler et al. [32], typically involve constructing a reward model from human preferences and subsequently optimizing the language model using specific algorithms such as Proximal Policy Optimization (PPO) [33]. These methods have been widely adopted for fine-tuning large language models [34–36], but they present significant challenges in terms of computational complexity and unstable training dynamics. The need to construct explicit reward models [37] adds further inefficiencies, particularly when human feedback involves nuanced judgments such as fluency, coherence, and factual accuracy. Recent approaches, like those in [36,38], emphasize the importance of leveraging human preference data for improving model alignment, but they continue to rely on the RLHF paradigm with explicit reward modeling.

DPO simplifies this process by eliminating the need for reward models and RL. Instead, it directly optimizes the policy of the language model using pairwise human preference data, building on the Bradley-Terry model [37]. This approach shares commonalities with preference-based ranking methods used in information retrieval [39] and recommendation systems [40], which also rely on pairwise comparisons. By formulating the optimization objective as a maximum likelihood problem over human preferences, DPO sidesteps the computational burdens of traditional RLHF methods, offering a more efficient and scalable solution. Similar efforts to simplify policy optimization via reward reparameterization, as explored by Peters and Schaal [41] and Korbak et al. [42], further support the design of DPO. Ultimately, DPO presents a streamlined and effective alternative to large-scale language models for keeping up with human preferences.

3. Method

3.1. Model Architecture

Our approach consists of three key steps: integrating human preference data, optimizing the model based on these preferences, and fine-tuning the policy, as illustrated

in Figure 1 (depicted on the right). Initially, we start with a model that's already been trained, such as GPT-3, and refine it through supervised learning on a task-specific dataset. DPO is introduced here as a more efficient alternative to traditional preference-based RL methods [32]. Unlike the conventional approach (shown on the left side of Figure 1), which requires training a reward model followed by RL, DPO (depicted on the right) bypasses the RL step altogether, directly optimizing the model using preference data. This reduces the complexity of the pipeline, eliminates the need for complex reward-based training, and results in faster and more scalable model fine-tuning.

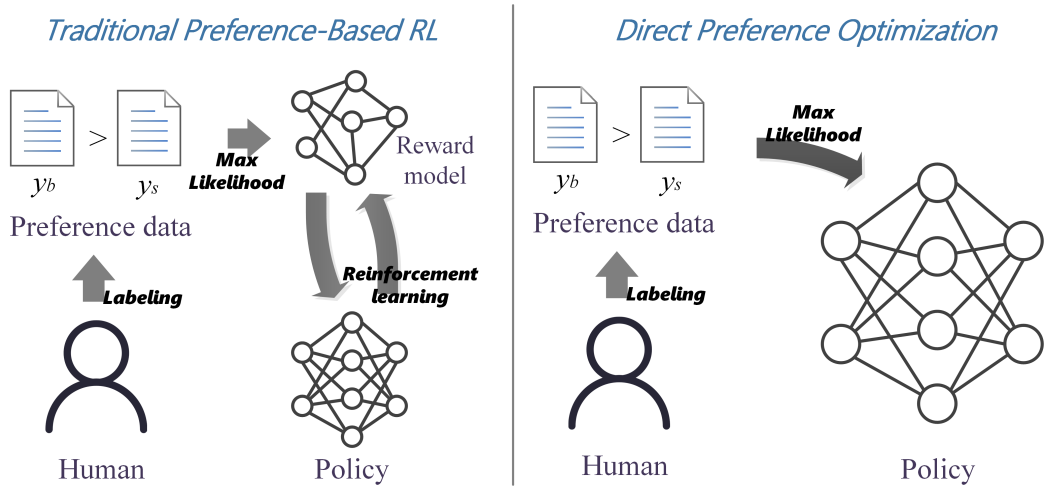


Figure 1. A comparison between traditional preference-based RL (**left**) and Direct Preference Optimization (DPO) (**right**).

3.2. Aligning with Human Judgement via Preference Data

The standard process entails refining a pre-trained language model using supervised learning on high-quality data tailored to the specific downstream task. However, this supervised fine-tuning (SFT) approach, while effective at generating grammatically correct and coherent responses, has a fundamental flaw: it optimizes for log-likelihood rather than aligning with human preferences. The log-likelihood objective encourages the model to assign probability mass to every response in the dataset, treating all responses as equally viable. This approach fails to distinguish between responses that have major errors, such as factual inaccuracies, and responses that are merely stylistically suboptimal. Consequently, even after extensive fine-tuning, the language model may produce low-quality outputs from the perspective of human evaluators.

To address this misalignment between the objective of the model and human evaluation criteria, DPO introduces human feedback directly into the fine-tuning process. Specifically, human preference data is collected by comparing pairs of responses to the same prompt. Given two responses, y_b and y_s, for a prompt x, where y_b is preferred over y_s, this preference data provides a more direct signal of what constitutes a high-quality response. Instead of merely maximizing the likelihood of responses that occurred in the dataset, DPO uses this preference data to optimize the language model towards generating responses that align with human judgment. The fine-tuning process no longer relies solely on the original dataset but incorporates these human judgments to drive the model towards better performance on real-world tasks.

3.3. Eliminating the Need for Explicit Reward Models

Another significant limitation of traditional preference-based RL methods is the need to construct an explicit reward model that gives a scalar reward to each response based on

its quality relative to human preferences. This reward model is usually trained on human-labeled data and subsequently used to guide the optimization of the language model via RL. The reward function $r(x, y)$, which takes a prompt x and its corresponding response y, is learned by fitting the reward model to human preferences, where the probability that response y_b is preferred over y_s is modeled using a Bradley-Terry framework:

$$p^*(y_b \succ y_s \mid x) = \frac{\exp(r(x, y_b))}{\exp(r(x, y_b)) + \exp(r(x, y_s))}. \tag{1}$$

This equation denotes the likelihood of one response being preferred over another based on their associated rewards. The challenge here, however, is that training an explicit reward model can be computationally expensive and often fails to capture the full complexity of human judgments, especially in tasks like summarization or dialogue generation, where multiple dimensions such as fluency, coherence, and factual accuracy need to be considered.

3.4. Simplifying Policy Optimization Without Reinforcement Learning

In this work, we resolve this problem by completely removing the requirement for an explicit reward model. In traditional RL-based models, during the RL phase, the learned reward function is used to provide feedback to the language model. Specifically, it formulates the following optimization problem:

$$\max_{\phi_\theta} \mathbb{E}_{x\sim\mathcal{D}, y\sim\phi_\theta(y|x)} \Big[r(x, y) \Big] - \beta\, D_{\mathrm{KL}}[\phi_\theta(y \mid x) \| \phi_{\mathrm{SFT}}(y \mid x)], \tag{2}$$

where β serves as a parameter to control the deviation from the base reference policy ϕ_{ref}, which corresponds to the initial SFT model ϕ_{SFT}. In practice, the language model policy ϕ_θ is similarly initialized to ϕ_{SFT}. For instance, a larger value of β reduces the deviation, making ϕ_θ more closely aligned with ϕ_{ref}, while a smaller β allows for greater divergence between the two. D_{KL} represents the Kullback-Leibler divergence [43]. $\mathbb{E}_{x\sim\mathcal{D}, y\sim\phi_\theta(y|x)}$ represents the expectation over the (x, y) pairs, where each sample x is drawn from the dataset $\mathcal{D}$, and the corresponding y is predicted by the model ϕ_θ given x. Rather than learning a separate reward function and then using RL to optimize the model, DPO model directly optimizes the language model based on human preferences. This is accomplished by leveraging a key insight from the RL literature: the optimal policy for a given reward function can be analytically represented using both the reward and the reference model (the fine-tuned supervised model). Specifically, the optimal policy under a given reward function is:

$$\phi_r(y \mid x) = \frac{1}{Z(x)} \phi_{\mathrm{SFT}}(y \mid x) \exp\left(\frac{1}{\beta} r(x, y)\right), \tag{3}$$

where $Z(x)$ denotes the partition function that ensures the policy sums to one:

$$Z(x) = \sum_y \phi_{\mathrm{SFT}}(y \mid x) \exp\left(\frac{1}{\beta} r(x, y)\right), \tag{4}$$

While this formulation simplifies the optimization problem by expressing the optimal policy directly in terms of the reward, it introduces a new challenge: computing the partition function $Z(x)$ is generally intractable, as it requires summing over all possible responses y.

DPO sidesteps this computational bottleneck by focusing on the relative difference in rewards between two responses rather than their absolute values. The Bradley-Terry model is determined only by the difference between the rewards for the two responses, y_b and y_s. By expressing the reward as:

$$r(x, y) = \beta \log \frac{\phi_r(y \mid x)}{\phi_{\text{SFT}}(y \mid x)} + \beta \log Z(x), \tag{5}$$

and substituting this into the Bradley-Terry preference model, the partition function $Z(x)$ cancels out, leading to the following expression for the preference probability:

$$p^*(y_b \succ y_s \mid x) = \sigma\left(\beta \log \frac{\phi_r(y_b \mid x)}{\phi_{\text{SFT}}(y_b \mid x)} - \beta \log \frac{\phi_r(y_s \mid x)}{\phi_{\text{SFT}}(y_s \mid x)}\right), \tag{6}$$

where $\sigma(x) = \frac{1}{1+\exp(-x)}$ denotes the sigmoid function. This formulation allows DPO to directly optimize the policy without computing the intractable partition function, focusing solely on the relative rewards of competing responses.

3.5. Direct Optimization of Human Preferences

In traditional RL-based methods, once the reward model is learned, RL techniques such as proximal policy optimization are used to optimize the policy by maximizing the expected reward of generated responses while penalizing deviations from the reference model. Although the RL objective in Equation (2) is effective, this approach significantly increases complexity. RL algorithms like PPO are computationally demanding, require careful tuning, and often suffer from high variance during training.

DPO simplifies the process by replacing RL algorithms with a maximum likelihood objective. Given a dataset $\hat{\mathcal{D}}$ of preference pairs (y_b, y_s) for prompts x, The goal of the DPO objective is to minimize the negative log-likelihood about the data represented by human preference:

$$\mathcal{L}_{\text{DPO}}(\boldsymbol{\theta}) = -\mathbb{E}_{(x,y_b,y_s)\sim\hat{\mathcal{D}}}\left[\log \sigma\left(\beta \log \frac{\phi(y_b \mid x)}{\phi_{\text{SFT}}(y_b \mid x)} - \beta \log \frac{\phi(y_s \mid x)}{\phi_{\text{SFT}}(y_s \mid x)}\right)\right]. \tag{7}$$

where expression $\mathbb{E}_{(x,y_b,y_s)\sim\hat{\mathcal{D}}}$ refers to the fact that the sample pairs (x, y_b, y_s) are drawn from the dataset $\hat{\mathcal{D}}$. This objective ensures that the language model is directly optimized to prefer responses that are more likely to be judged favorably by humans. The DPO loss function operates on the relative difference between the rewards assigned to the preferred response $\boldsymbol{y}_b$ and the dispreferred response $\boldsymbol{y}_s$, rather than their absolute values.

3.6. Gradient-Based Update for Fine-Tuning

The gradient of the DPO objective concerning the model parameters $\boldsymbol{\theta}$ is expressed as:

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\text{DPO}} = -\beta \mathbb{E}_{(x,y_b,y_s)\sim\hat{\mathcal{D}}}\Big[\sigma(r_{\boldsymbol{\theta}}(x, y_s) - r_{\boldsymbol{\theta}}(x, y_b))(\nabla_{\boldsymbol{\theta}} \log \phi(y_b \mid x) - \nabla_{\boldsymbol{\theta}} \log \phi(y_s \mid x))\Big]. \tag{8}$$

The gradient formula is intended to give a intuitive understanding of what happens during the gradient descent process. This gradient drives the model to raise the likelihood of generating preferred responses while reducing the likelihood of generating non-preferred ones, based on the implicit reward function $r_{\boldsymbol{\theta}}(x, y)$ derived from the policy.

In conclusion, the model DPO significantly simplifies the process of fine-tuning large-scale language models based on human preferences. By optimizing the model directly using preference data, without using explicit reward models or RL, we can fine-tune the large language model to better meet our requirements.

4. Experiments

4.1. Datasets

We fine-tuned the GPT-3 model with 175 Billion parameters specifically for generating and evaluating poetry using a widely accessible Romanian poetry dataset. Furthermore,

we collected a novel dataset, employing it as the foundation for our quality evaluation system. Our dataset comprises modern Chinese renditions [44,45] of Romanian poetry, spanning a total of 79,000 characters, each meticulously vetted and refined to ensure accuracy. The collection of poems within the dataset spans seven distinct thematic categories. Each poem has been meticulously tagged on a word-level basis across 22 grammatical classifications, aligning with the lexical structures of both Chinese and Romanian languages. The distribution of poetry types in the dataset is illustrated in Table 1.

Table 1. Distribution of poetry types in the dataset.

Dataset	Types of Poetry	Quantity	Percentage
Modern Chinese versions of the works of Romanian poets (79,000 Chinese characters, 328 poems)	Lyric Poetry	65	19.8%
	Love Poetry	104	31.7%
	Narrative Poetry	39	11.9%
	Philosophical Poetry	33	10.1%
	Pastoral Poetry	46	14.0%
	Satirical Poetry	15	4.6%
	Mourning Poetry	26	7.9%

4.2. Collecting Data for Evaluation

Our dataset is primarily composed of text prompts submitted to the GPT API (Application Programming Interface), which is a set of protocols and tools that allows different software applications to communicate with each other. Generally, these prompts convey the task directly using plain language instructions (e.g., "Compose a Chinese philosophical poem in the style of Romanian literature"). However, the prompts may also define the task in more implicit ways, such as through partial continuations like starting a Chinese rendition of a Romanian poem.

To develop our comparative dataset and carry out a preliminary evaluation, we established a team of ten contractual annotators, all of whom are native Chinese speakers with proficient understanding of poetry and extensive familiarity with Romanian literary works, equipped to precisely distinguish the Romanian style. In order to effectively incorporate Romanian style in model training, the annotations take into account both Chinese syntactic structures and specific structures in Romanian. Annotators work hard to capture the user's intent contained within the prompts and evaluate the results according to their background knowledge and the guidelines we provide, as shown below.

Guidelines: Annotators are requested to select their favorite poem by considering the following criteria: relevance to the theme, coherence of the content, originality, and the use of form, vocabulary, syntax, figurative language, idioms, and cultural references.

5. Results and Evaluation

Some samples of verses generated by the baseline model and the fine-tuned model are shown in Table 2. It can be seen from Table 2 that the poems generated by the fine-tuned model have more Romanian poetic images and expressions than those generated by the baseline model, for example: "蒂米什瓦拉 (Timisoara)", "城堡 (castle)". In addition, the description of the poems generated by the fine-tuned model are also relatively more delicate, for example: "他总是说自己是诗人，虽然没有人读他那沉重而冰冷的诗歌 (He always claimed to be a poet, even though no one read his heavy and icy poems)". The two verses above clearly outlines the image of a lonely and melancholic poet who is unremittingly pursuing art.

Table 2. Samples generated poems and their corresponding translations.

Model	Generated Verses	Generated Verses in English
Original Pre-trained Model	所以我会继续前行， 为我的梦想和理想而战， 不被困难和阻碍打败， 向世界展示像我这样的人存在。	*So I will keep going,* *Fight for my dreams and ideals,* *I will not be defeated by difficulties and obstacles,* *Show the world that people like me exist.*
	我喜欢看雨落在 松树叶上， 感觉自己就像是唯一的存在， 在这个广阔的世界里。	*I like to watch the rain fall* *On the pine leaves,* *Feel like I am the only being,* *In this vast world.*
	我记得以前的日子， 我的心跳得更快， 太阳在天空的一角升起， 照耀着绿色的田野和鲜花。	*I remember the old days,* *My heart used to beat faster,* *The sun rises in the corner of the sky,* *Shining on the green fields and flowers.*
	我渴望自由， 如同鸟儿飞翔， 却又被命运所束缚， 无法挣脱这束缚。	*I yearn for freedom,* *Flying like a bird,* *But I am bound by fate,* *Unable to break free from this shackle.*
	现在，当我看着高高挂在天空中的月亮， 星星似乎在向我微笑， 我感到我已经找到了我所寻求的， 我的生命有一个神圣的意义。	*Now, when I look at the moon high in the sky,* *The stars seem to be smiling at me,* *I feel I have found what I am looking for,* *My life has a sacred meaning.*
Fine-tuned Pre-trained Model	在蒂米什瓦拉的街道上， 你是我见过最美的风景， 你的名字如同古老的城堡， 在我心中徜徉。	*On the streets of Timisoara,* *Thou art the most beautiful scenery I have ever seen,* *Thy name, like an ancient castle,* *Roams in my heart.*
	他颤抖着手，拿着一本旧笔记本， 里面写着匆忙而又被擦掉的诗句， 他总是说自己是诗人，虽然 没有人读他那沉重而冰冷的诗歌。	*He trembled as he held an old notebook,* *Inside were hastily written and erased verses,* *He always claimed to be a poet, even though* *No one read his heavy and icy poems.*
	闭着疼痛的眼睛， 感受时间像一条河流流逝， 在黑暗而沉重的夜晚里， 被遗忘的回忆再次重生。	*With my eyes closed in pain,* *I feel time flowing like a river,* *In the dark and heavy night,* *Forgotten memories are reborn again.*
	在早晨的阳光中， 阳台上的花儿在微风中摇曳， 叶子上的露珠闪耀着光芒， 如同千万颗钻石在跳跃。	*In the morning sunshine,* *The flowers on the balcony sway in the breeze,* *Dewdrops on the leaves sparkling,* *Just like millions of diamonds dancing.*
	我们站在时间的边缘， 寻找意义， 但也许生命只是一场游戏， 而我们只是舞台上的演员。	*We stand on the edge of time,* *Looking for meaning,* *But maybe life is just a game,* *And we are just actors on the stage.*

Since the existing machine evaluation methods cannot accurately reflect the quality of the generated stylized poems, we employ the following two approaches to assess the generated results: human evaluation and statistics on poetry-image coincidence.

5.1. Human Evaluation

We invited 26 evaluators who are familiar with European literature and have good appreciation ability to independently score 35 poems. The average length of the 35 poems is 20.17 lines, with a mean of 9.62 characters for each line. To ensure both objectivity and effectiveness in the evaluation process, the poems assessed by the evaluators included the Chinese versions of the real Romanian poets' works as a reference group in addition to the generated poems. To ensure a comprehensive evaluation, we compared our model with the state-ofthe-art poetry generation method, SPG [46]. This advanced method incorporates mutual information to enhance the fluency and coherence of generated poems. The 35 poems for evaluation include 10 poems created by the pre-trained model without further

tuning, 10 poems produced after training, 10 poems produced by SPG and 5 translations from works by well-known Romanian poets (we selected poems that were are relatively unknown in China to make sure that evaluators were unfamiliar with them previously). These 35 poems are arranged in random order and do not contain any information about the author or title.

Since the generated poetry involves two languages and cultural backgrounds, we did not use the four-aspect evaluation approach [17,22] (I. Fluency II. Coherence III. Meaning IV. Poeticness) and the six-aspect evaluation approach [13,26] (I. How representative is the text of poetry? II. How clear and comprehensible is it? III. How polished and effective is the linguistic use? IV. Can the text create mental picture? V. Does the text provoke feelings or emotions? VI. How much does the subject like the text?) commonly used in the evaluation of monolingual poetry generation in the past, but evaluated from seven more specific dimensions: grammar, logicality, rhetoric creativity, depth, style and human-likeness. Among them, "grammar" and "logicality" are basic perspectives, which test the readability of the poem; "rhetoric", "creativity" and "depth" are advanced perspectives, which test the level of artistic creation of the poem; "style" evaluates the degree of realization of the poetry in the Romanian style. In addition, "human-likeness" is also added to the rating scale. There are a thousand way to interpret a poem, in terms of evaluation in the field of art, affected by cultural background, educational background, personal experience, aesthetic preferences, etc., there may be great differences in the feelings of the same work. Poetry creation is open, and when the basic grammatical logic reaches a normal level, there is no fixed answer to compare. We believe that for the generation of stylized poetry, the most important thing is not to conform to everyone's aesthetics to the greatest extent, but to be "more human-like". Therefore, the dimension of human-likeness was added to the evaluation. A poem written by humans will make readers feel very natural. Fluent language, contextual relevance, and consistent reference are all important factors in determining whether a poem is "human-like". Therefore, poems with problems such as repetitive language, irrelevant context, and inconsistent references will be judged as less human-like poems.

The rating for each viewpoint is displayed on a 5-point Likert Scale, and the detailed evaluation criteria are outlined in Table 3.

Table 3. Assessment Aspects and Judging Standards.

Grammar	Evaluation content	Fluency and grammatical correctness of the poetry.
	Evaluation criteria	5 Excellent; 4 Good; 3 Average; 2 Below Average; 1 Poor.
Logicality	Evaluation content	Clarity of thinking, contextual coherence, lack of inconsistency.
	Evaluation criteria	5 Excellent; 4 Good; 3 Average; 2 Below Average; 1 Poor.
Rhetoric	Evaluation content	Language richness and rhetorical appropriateness.
	Evaluation criteria	5 Excellent; 4 Good; 3 Average; 2 Below Average; 1 Poor.
Creativity	Evaluation content	Uniqueness of expression, novelty of imagery, no clichés.
	Evaluation criteria	5 Excellent; 4 Good; 3 Average; 2 Below Average; 1 Poor.
Depth	Evaluation content	Inspiring and thought-provoking.
	Evaluation criteria	5 Excellent; 4 Good; 3 Average; 2 Below Average; 1 Poor.
Style	Evaluation content	Similarity to Romanian poetry style.
	Evaluation criteria	5 Very High; 4 High; 3 Average; 2 Below Average; 1 Very Low.
Human-likeness	Evaluation content	Likelihood of the poetry written by human.
	Evaluation criteria	5 Definitely; 4 Very Probably; 3 Probably; 2 Probably Not; 1 Definitely Not.

A total of 6370 scoring values were collected, and Figure 2 displays the average values for each dimension: Human-written indicates the score of works created by humans, PTM and FTM denotes the score of the poems produced by the pre-trained and the fine-tuned models, respectively.

Cronbach's α and McDonald's ω are two coefficients suitable for evaluating the reliability of Likert scale results. Cronbach's α and McDonald's ω both have values above 0.8, reflecting very high reliability; values between 0.7 and 0.8 indicate that high reliability. Table 4 shows that the calculated results of Cronbach's α and McDonald's ω are all over 0.7, indicating that the human evaluation data is reliable.

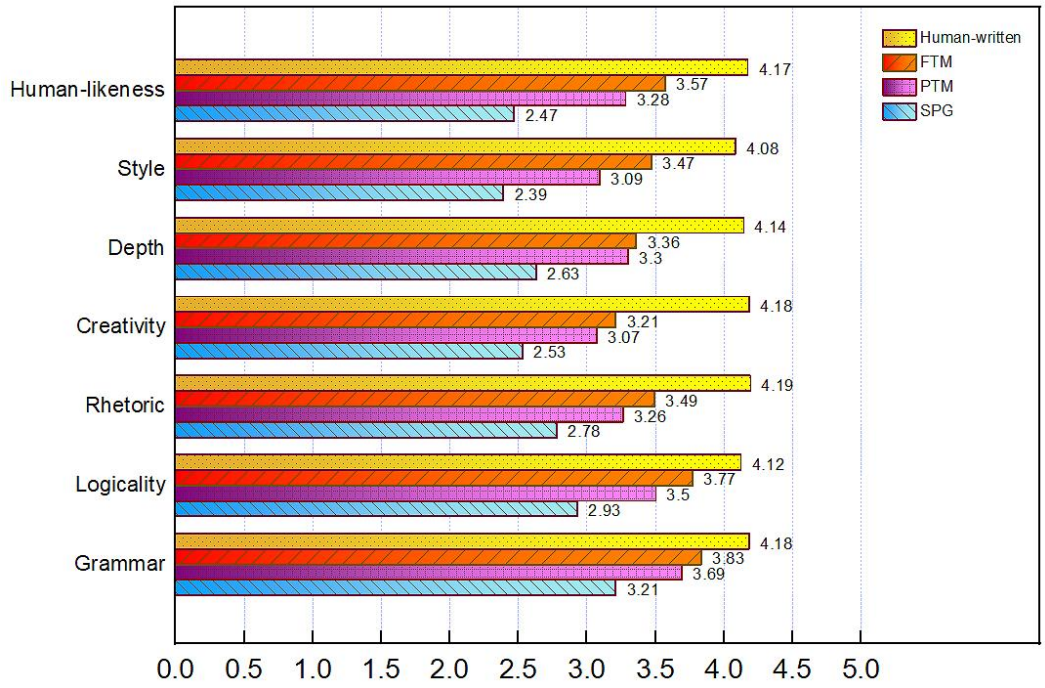


Figure 2. Human evaluation results of original pre-trained model-produced poems, DPO model-produced poems, SPG method produced poems and human-written poems.

Human evaluation results show that the poems produced by the original pre-trained and fine-tuned models are above-average (i.e., average score > 3) in all evaluation dimensions. The model trained by DPO outperforms the original pre-trained model in each dimension: 0.14 points in grammar, 0.27 points in logicality, 0.23 points in rhetoric, 0.14 points in creativity, 0.06 points in depth, 0.38 points in style and 0.29 points in human-likeness. It can be seen that DPO is of great help to the logicality, the rhetoric, and the human-likeness of generated poetry, especially to the improvement of the proximity to the target style. The DPO-trained models achieved 91.6%, 91.5%, 83.33%, 76.8%, 76.2%, 85.1%, and 85.6% of the scores of works produced by real poets in terms of grammar, logicality, rhetoric, creativity, depth, style, and human-likeness, respectively. It can be seen that the scores of generated poems in terms of grammar and logicality are significantly higher than those in other dimensions, reaching more than 90% of the scores of real poets' works, which is relatively closer to the performance of poets' works in these two dimensions. However, in terms of creativity and depth, there is still a big gap between the generated poems and the poems of excellent poets.

Table 4. Trustworthiness of human evaluation outcomes.

Reliability	Grammar	Logicality	Rhetoric	Creativity	Depth	Style	Human-Likeness
Cronbach's α	0.882	0.845	0.878	0.899	0.868	0.807	0.859
McDonald's ω	0.894	0.865	0.885	0.896	0.882	0.756	0.870

5.2. Imagery Evaluation

"Imagery" carries distinct cultural and regional characteristics, and playing a crucial role in shaping the style of poetry. The commonly used poetic images in poetry in different languages are different. For example, the common plant "竹 (bamboo)" in China can refer to a man of moral integrity in Chinese poetry; "Feţi-frumoşi" is a character in Romanian folk tales, which can refer to a handsome man in poetry. "麒麟 (Kylin)" is an animal in Chinese mythology, which is a symbol of luck and auspiciousness, while "Zburător" is the spirit who torments girls' sleep in Romanian folk mythology, and is also the idealized embodiment of the lover in Romanian literature. The same poetic image may have different connotations in different cultural backgrounds, and sometimes even the same poetic image in the same language may be associated with opposite emotions in different contexts. For instance, which can be seen in Table 5, the poem "Înger şi demon" by the famous Romanian poet Mihai Eminescu is titled "Angel and Demon", but the "angel" and the "demon" here do not refer to sacred and evil mythological creatures, but as the poetic images respectively refer to the kind and gentle princess, and the brave, idealistic and rebellious youth who is regarded as a "demon" by the ruler in the poem.

Table 5. Example of poetic images in Romanian poetry.

Verses from "Înger şi demon"	Translation in English
Ea un înger ce se roagă - El un demon ce visează;	"She" a praying angel—"He" a dreaming demon;
Ea o inimă de aur—El un suflet apostat;	"She" a heart of gold—"He" an apostate soul;
El , în umbra lui fatală, stă-ndărătnic răzemat—	"He", in his fatal shadow, stubbornly leans—
La picioarele Madonei, tristă, sfântă, Ea veghează.	At the feet of the Madonna, sad, holy, She watches.

Because of the importance of poetic imagery to poetic style, we assess the quality of the produced poems by analyzing the distribution of imagery. Firstly, we counted the 50 most frequently occurring images in 200 representative works of 80 most famous modern poets in China and 330 representative works of 46 famous poets in Romania. Then, we generated 50 poems through the original pre-training model, and 50 poems were also generated through the model trained by DPO, and counted the 50 most frequently occurring poetic images of each group. Finally, we calculated the coincidence rate with the top 10, the top 20, the top 30, the top 40 and the top 50 high-frequency poetic images between Romanian poems and Chinese poems, between Romanian poems and original pre-trained model-generated poems, as well as between Romanian poems and DPO-trained model-generated poems. The statistical results of coincidence rate are shown in the Figure 3: CR represents the coincidence rate of poetic imagery between Romanian poetry and Chinese modern poetry, GR1 represents the coincidence rate of poetic images between pre-trained model generated poetry and Romanian poetry, GR2 represents the coincidence rate of poetic images between pre-trained model with DPO generated poetry and Romanian poetry.

Figure 3 shows that the overlap rate of shared imagery between Chinese and Romanian poetry is relatively low, falling below 45%. The model trained by DPO has the same coincidence rate with Romanian poetry on the top 10 high-frequency poetic images as the baseline model; its coincidence rate with Romanian poetry in the top 20, top 30, top 40, and top 50 high-frequency poetic images is 20%, 13.3%, 12.5%, and 8% higher than the baseline model, respectively. The fine-tuned model improves by 0.38 points in the dimension of style

in human evaluation, which is the most improved among all seven dimensions, combined with the statistical results of the distribution of poetic images, it shows that DPO training can effectively improve the proximity of generated poetry to Romanian style.

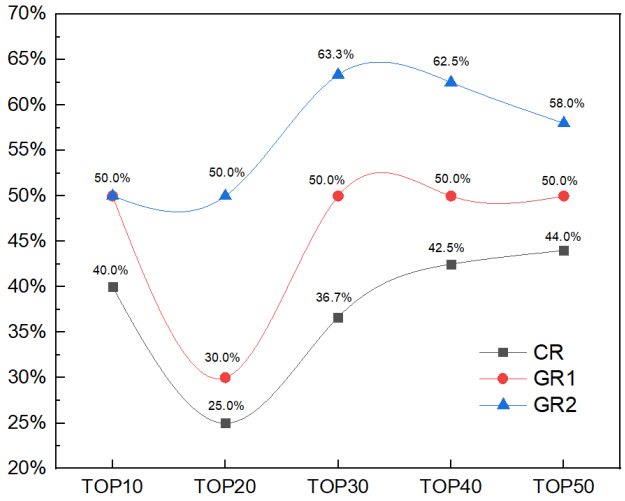


Figure 3. Distribution of high-frequency poetic images in Romanian poems, Chinese poems and DPO model-generated poems.

Figure 4a–d offers a detailed and visual representation of the distribution of high-frequency poetic imagery in Chinese poetry, Romanian poetry, and the Romanian-style Chinese poetry generated by the DPO-trained model.

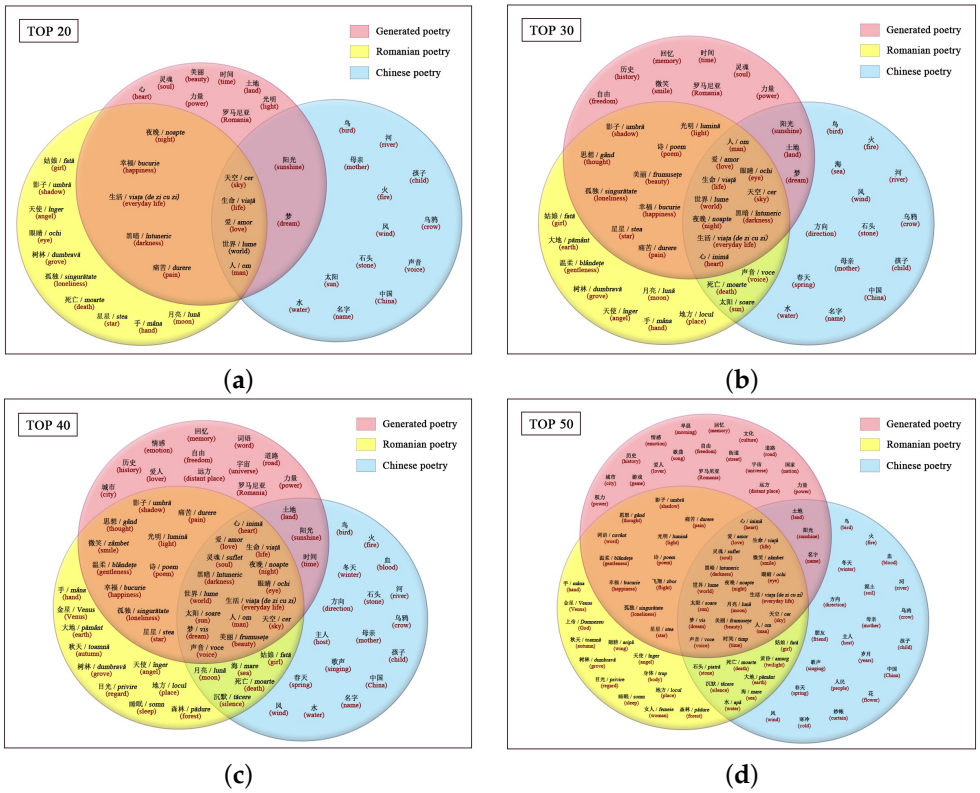


Figure 4. Comparison of common poetic imagery found in Romanian poetry, Chinese poetry, and model-generated verses. (**a**) Analysis of the top 20 most frequently occurring poetic images. (**b**) Analysis of the top 30 most frequently occurring poetic images. (**c**) Analysis of the top 40 most frequently occurring poetic images. (**d**) Analysis of the top 50 most frequently occurring poetic images.

6. Conclusions

In this paper, we presented a method for the automatic generation of modern Chinese poetry in the Romanian style. The approach involved fine-tuning a pre-trained model as the baseline, followed by applying Direct Preference Optimization (DPO) to improve the quality and stylistic elements of the generated poems. Human evaluators rated these poems on a 5-point scale across seven criteria—grammar, logicality, rhetoric, creativity, depth, style, and human-likeness—achieving scores between 3.21 and 3.83. This corresponds to 76.2% to 91.6% of the scores given to Chinese translations of authentic Romanian poetry. Additionally, the overlap of the 30th to 50th most frequently occurring poetic images between the generated and authentic Romanian poems ranged from 58.0% to 63.3%. These results suggest that DPO significantly enhances the stylization process, enabling the model to produce modern Chinese poems with a distinct Romanian influence. However, while the model performed well in fundamental aspects such as grammar and logicality, it still lagged behind real poets in areas like creativity and depth. Future work will focus on refining the model to address these limitations and further enhance its poetic capabilities.

Author Contributions: Conceptualization, L.Z.; Methodology, L.Z., D.Z. and Y.Z.; Validation, L.Z., D.Z. and Y.Z.; Formal analysis, L.Z.; Investigation, L.Z.; Data curation, L.Z.; Writing—original draft, L.Z.; Writing—review & editing, L.Z., D.Z., Y.Z. and G.W.; Visualization, L.Z., D.Z. and Y.Z.; Supervision, G.W.; Funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by the Educational Department of Liaoning Province in China under scientific research project number LJKR0002.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gervas, P. WASP: Evaluation of different strategies for the automatic generation of Spanish verse. In Proceedings of the AISB00 Symposium on Creative & Cultural Aspects of AI, Birminham, UK, 17–20 April 2000; pp. 93–100.
2. Oliveira, H.G. PoeTryMe: A versatile platform for poetry generation. In Proceedings of the Computational Creativity, Concept Invention, and General Intelligence (C3GI), Montpellier, France, 27 August 2012; pp. 21–26.
3. Jiang, L.; Zhou, M. Generating Chinese couplets using a statistical MT approach. In Proceedings of the 22nd International Conference on Computational Linguistics, Manchester, UK, 18–22 August 2008; pp. 377–384.
4. He, J.; Zhou, M.; Jiang, L. Generating Chinese classical poems with statistical machine translation models. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; Volume 26, pp. 1650–1656.
5. Wang, L.; Yu, Y.; Zhang, Y. Custom Generation of Poetry Based on Seq2Seq Model. *J. Front. Comput. Sci. Technol.* **2020**, *14*, 1028–1035.
6. Hopkins, J.; Kiela, D. Automatically generating rhythmic verse with neural networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 168–178.
7. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 4171–4186.
8. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
9. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
10. Hämäläinen, M.; Alnajjar, K.; Poibeau, T. Modern French poetry generation with RoBERTa and GPT-2. In Proceedings of the 13th International Conference on Computational Creativity (ICCC), Bolzano, Italy, 27 June–1 July 2022; pp. 12–16.
11. Nehal, E.; Mervat, A.; Maryam, E.; Mohamed, A. Generating Classical Arabic Poetry using Pre-trained Models. In Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP), Abu Dhabi, United Arab Emirates, 8 December 2022; pp. 53–62.

12. Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C.D.; Ermon, S.; Finn, C. Direct preference optimization: Your language model is secretly a reward model. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023.
13. Shihadeh, J.; Ackerman, M. EMILY: An Emily Dickinson machine. In Proceedings of the 11th International Conference on Computational Creativity (ICCC'20), Coimbra, Portugal, 7–11 September 2020; pp. 243–246.
14. Ghazvininejad, M.; Shi, X.; Choi, Y.; Knight, K. Generating topical poetry. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 1183–1191.
15. Li, J.; Song, Y.; Zhang, H.; Chen, D.; Shi, S.; Zhao, D.; Yan, R. Generating classical Chinese poems via conditional variational autoencoder and adversarial training. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 31 October–4 November 2018; pp. 3890–3900.
16. Gao, T.; Xiong, P.; Shen, J. A new automatic Chinese poetry generation model based on neural network. In Proceedings of the 2020 IEEE World Congress on Services (SERVICES), Beijing, China, 18–23 October 2020; pp. 41–44.
17. Zhang, X.; Lapata, M. Chinese poetry generation with recurrent neural networks. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 670–680.
18. Zhang, H.; Zhang, Z. Automatic generation method of ancient poetry based on LSTM. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9–13 November 2020; pp. 95–99.
19. Liu, Z.; Fu, Z.; Cao, J.; De Melo, G.; Tam, Y.C.; Niu, C.; Zhou, J. Rhetorically controlled encoder-decoder for modern Chinese poetry generation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1992–2001.
20. Hirata, K.; Yokoyama, S.; Yamashita, T.; Kawamura, H. Implementation of autoregressive language models for generation of seasonal fixed-form Haiku in Japanese. *IIAI Lett. Inform. Interdiscip. Res.* **2023**, *3*, LIIR075. [CrossRef]
21. Shao, G.; Kobayashi, Y.; Kishigami, J. Traditional Japanese Haiku generator using RNN language model. In Proceedings of the 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 9–12 October 2018; pp. 263–264.
22. Talafha, S.; Rekabdar, B. Arabic poem generation with hierarchical recurrent attentional network. In Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 30 January–1 February 2019; pp. 316–323.
23. Gonçalo Oliveira, H.; Hervás, R.; Díaz, A.; Gervás, P. Adapting a generic platform for poetry generation to produce Spanish poems. In Proceedings of the 5th International Conference on Computational Creativity (ICCC), Ljubljana, Slovenia, 9–13 June 2014; pp. 63–71.
24. Gonçalo Oliveira, H.G.; Cardoso, A. Poetry generation with PoeTryMe. In *Computational Creativity Research: Towards Creative Machines*; Besold, T., Schorlemmer, M., Smaill, A., Eds.; Atlantis Press: Amsterdam, The Netherlands, 2015; Volume 7, pp. 243–266.
25. Hämäläinen, M.; Alnajjar, K. Let's FACE it. Finnish poetry generation with aesthetics and framing. *arXiv* **2019**, arXiv:1910.13946.
26. Toivanen, J.M.; Toivonen, H.; Valitutti, A.; Gross, O. Corpus-based generation of content and form in poetry. In Proceedings of the 3rd International Conference on Computational Creativity (ICCC), Dublin, Ireland, 30 May–1 June 2012; pp. 211–215.
27. Das, A.; Gambäck, B. Poetic Machine: Computational creativity for automatic poetry generation in Bengali. In Proceedings of the International Conference on Innovative Computing and Cloud Computing, Guilin, China, 19–21 October 2014; pp. 230–238.
28. Ahmad, S.; Joglekar, P. Urdu and Hindi poetry generation using neural networks. In *Data Management, Analytics and Innovation, Proceedings of the ICDMAI 2022, Kolkata, India, 17–19 January 2025*; Goswami, S., Barara, I.S., Goje, A., Mohan, C., Bruckstein, A.M., Eds.; Lecture Notes on Data Engineering and Communications Technologies Series; Springer: Singapore, 2023; Volume 137.
29. Zhang, L.; Zhang, R.; Mao, X.; Chang, Y. QiuNiu: A Chinese lyrics generation system with passage-level input. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics System Demonstrations, Dublin, Ireland, 22–27 May 2022; pp. 76–82.
30. Liu, N.; Han, W.; Liu, G.; Peng, D.; Zhang, R.; Wang, X.; Ruan, H. ChipSong: A controllable lyric generation system for Chinese popular song. In Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), Dublin, Ireland, 26 May 2022; pp. 85–95.
31. Zhang, R.; Mao, X.; Li, L.; Jiang, L.; Chen, L.; Hu, Z.; Xi, Y.; Fan, C.; Huang, M. Youling: An AI-assisted lyrics creation system. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Online, 16–20 November 2020; pp. 85–91.
32. Ziegler, D.M.; Stiennon, N.; Wu, J.; Brown, T.B.; Radford, A.; Amodei, D.; Christiano, P.; Irving, G. Fine-tuning language models from human preferences. *arXiv* **2019**, arXiv:1909.08593.
33. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
34. Stiennon, N.; Ouyang, L.; Wu, J.; Ziegler, D.; Lowe, R.; Voss, C.; Radford, A.; Amodei, D.; Christiano, P.F. Learning to summarize with human feedback. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33, pp. 3008–3021.

35. Bai, Y.; Jones, A.; Ndousse, K.; Askell, A.; Chen, A.; DasSarma, N.; Drain, D.; Fort, S.; Ganguli, D.; Henighan, T.; et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv* **2022**, arXiv:2204.05862.
36. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 27730–27744.
37. Bradley, R.A.; Terry, M.E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **1952**, *39*, 324–345. [CrossRef]
38. Wu, J.; Ouyang, L.; Ziegler, D.M.; Stiennon, N.; Lowe, R.; Leike, J.; Christiano, P. Recursively summarizing books with human feedback. *arXiv* **2021**, arXiv:2109.10862.
39. Joachims, T. Optimizing search engines using clickthrough data. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002.
40. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. *arXiv* **2012**, arXiv:1205.2618.
41. Peters, J.; Schaal, S. Reinforcement learning by reward-weighted regression for operational space control. In Proceedings of the 24th International Conference on Machine Learning, Corvalis, OR, USA, 20–24 June 2007.
42. Korbak, T.; Elsahar, H.; Kruszewski, G.; Dymetman, M. On reinforcement learning and distribution matching for fine-tuning language models with no catastrophic forgetting. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 16203–16220.
43. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
44. Lucian, B.; Maria, B.; Stefan, A.D.; Nina, C.; Petre, S.; Nichita, S.; Marin, S.; Gheorghe, G.; Cezar, B.; Dan, L.; et al. *Anthology of Contemporary Romanian Lyric Poetry*; Gao, X., Translator; Flower City Press: Guangzhou, China, 2018.
45. Mihai, E. *Mihai Eminescu, Poezii*; Ding, C., Constantin, L., Eds.; Ge, B., Xu, W., Li, N., Feng, Z., Translator; Foreign Language Teaching and Research Press: Beijing, China, 2018.
46. Yang, C.; Sun, M.; Yi, X.; Li, W. Stylistic Chinese Poetry Generation via Unsupervised Style Disentanglement. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3960–3969.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.