

PatAlign: Bridging the Gap between Scientific Papers and Patent Descriptions via Iterative Preference Optimization

Zheming Fan^{1,2}, Xuemeng Song¹, Min Yang³, Linyi Yang^{1†}

¹Southern University of Science and Technology

²Shandong University

³Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

fanzheming@mail.sdu.edu.cn, sxmustc@gmail.com

min.yang@siat.ac.cn, yangly6@sustech.edu.cn

Abstract

Transforming academic papers into patent descriptions challenges LLMs with extreme length, strict factuality, and rigorous stylistic norms. Existing supervised methods suffer from pathological repetition, while static preference optimization causes style degradation by erroneously penalizing necessary structural repetitions. To address this, we propose PATALIGN, a framework utilizing Iterative Preference Optimization to dynamically sample negative instances from the evolving policy, effectively distinguishing pathological repetition from the structural repetition required to ensure patent rigor. We also introduce the Paper-to-Patent Score (P²S) metric to evaluate the overall quality of generated patents. Experiments demonstrate that PATALIGN effectively mitigates pathological repetition and style degradation, achieving state-of-the-art performance in both factual consistency and language style. Notably, our 8B model significantly outperforms 32B general-purpose models, validating the efficacy of iterative alignment strategies in complex long-form generation tasks. Our code and data are available at <https://github.com/SUSTech-GenAI/PatAlign>.

1 Introduction

Patent documents are an important form of intellectual property that protects the interests of individuals and organizations. However, unlike general web documents, patent documents are typically characterized by extreme length, a highly formalized writing style, and a dense concentration of technical terminology (Zhang et al., 2015). Consequently, manual patent drafting is a time-consuming task that requires extensive domain expertise. With the rapid development of large language models (LLMs) (Chang et al., 2024; Shukla et al., 2025), automated patent writing has entered a transformative era (Wang et al., 2024b; Trippe and Lee, 2025),

promising to significantly improve the productivity of patent practitioners by automatically generating patent documents (Murray and Stern, 2007).

Recent research on patent generation has made notable progress, with prior studies primarily focusing on brief components such as (1) preliminary conceptual texts (Ren et al., 2025); (2) patent abstracts or outlines (Codina-Filbà et al., 2017); (3) patent claims (Jiang et al., 2025a,b; LI et al., 2025; Lee, 2025); and (4) patent figure descriptions (Shukla et al., 2025). However, although patent descriptions encompass the core technical facts and account for over 90% of the entire patent document (Knappich et al., 2025), systematic research dedicated to the generation of full patent descriptions remains relatively limited.

Although a recent study, AutoPatent (Wang et al., 2024b), attempts to generate complete patent documents, it strictly relies on manually written patent drafts as input. In real-world scenarios, however, researchers more frequently need to directly convert scientific discoveries presented in academic papers into patents for intellectual property protection. To address this practical need, Knappich et al. (2025) attempted to utilize academic papers as ready-made invention disclosures to automatically generate full patent descriptions. Nevertheless, their approach relies on a train-free, chunk-based generation strategy. Given the significant discrepancy in linguistic style between academic papers and patents, such general-purpose base models fundamentally struggle to adapt to the highly formalized language of authentic patents.

Therefore, exploring an effective training framework for the paper-to-patent generation task is of paramount importance, but it also faces significant challenges: the generated long-form text must not only faithfully reflect the technical facts of the source paper but also strictly comply with the unique linguistic conventions of the patent domain. This imposes severe challenges on existing LLM

[†]Corresponding author <yangly6@sustech.edu.cn>.

training paradigms in terms of generation length, factual consistency, and style alignment. Given the inherent complexity and difficulty of the paper-to-patent generation task, current offline training paradigms exhibit significant structural limitations. Specifically, supervised fine-tuning (SFT) tends to induce severe pathological repetition because the model fails to distinguish between the legitimate, necessary structural repetition inherent in patent texts and degenerate pathological repetition, resulting in generation collapse (Knappich et al., 2025).

While Direct Preference Optimization (DPO) (Rafailov et al., 2023) can alleviate this issue by utilizing authentic patents as positive samples and pathologically repetitive texts as negative samples, these static negative samples fail to precisely align with the most prominent generation flaws of the continually evolving policy. Consequently, the persistent optimization focus on penalizing repetitive features in the later stages of training leads to an over-correction problem, which erroneously suppresses the legitimate structural repetitions required in patent texts.

To this end, we propose PATALIGN, a dynamic training framework based on iterative preference optimization (Pang et al., 2024; Yuan et al., 2024). PATALIGN constructs a curriculum of preference learning with progressively increasing difficulty: at each iteration, the model is optimized to distinguish reference patents, representing the true data distribution, from evolving negative samples generated by the current policy. In early iterations, this contrast primarily drives the model to suppress pronounced pathological repetition; in later iterations, as negative samples gradually approach real patents in factual consistency and linguistic style, the optimization focus shifts to subtle stylistic deviations, guiding the model to move attention from coarse repetition suppression to fine-grained style discrimination, fundamentally addressing the style degradation caused by static alignment paradigms.

We conduct extensive experiments across multiple dimensions, demonstrating that PATALIGN achieves superior performance on the paper-to-patent generation task compared to the prior state-of-the-art baseline (Knappich et al., 2025). Specifically, compared with COPGEN, PATALIGN improves the factual consistency metric (*Pap*→*Gen*) by 25.5 percentage points and the language style score by 20.9 percentage points. To comprehensively evaluate the generation quality, we propose the Paper-to-Patent Score (P²S) metric, on which

it achieves 82.7%, surpassing the COPGEN baseline by 23.5 percentage points. Compared with the SFT baseline, PATALIGN not only enhances factual consistency but also reduces the repetition rate ($RR > 80$) by 9.9 percentage points.

To our knowledge, we are the first to investigate the effectiveness of iterative preference optimization in the paper-to-patent task. Our main contributions are summarized as follows:

- We propose an effective training paradigm of LLMs on the paper-to-patent task, overcoming the pathological repetition problem.
- We introduce a novel composite evaluation metric, the Paper-to-Patent Score (P²S), for holistic assessment of paper-to-patent conversion quality.
- We establish a new state-of-the-art result in the paper-to-patent task in factual consistency, linguistic style, and overall quality.

2 Related Work

Patent Generation. The research paradigm of patent natural language processing (Patent NLP) has gradually evolved from traditional discriminative tasks—such as patent analysis, classification, and assisted retrieval (Tseng et al., 2007; Yoshikawa and Krestel, 2024; Casola and Lavelli, 2022)—to more complex generative applications. Early generative studies mainly focused on drafting specific sections, including patent abstract generation (Codina-Filbà et al., 2017; Trappey et al., 2020; Kim and Yoon, 2022) and claim generation (Lee and Hsiang, 2020; Jiang et al., 2025a,b; Li et al., 2025; Lee, 2025). Meanwhile, the research scope expanded to transformation and enhancement of internal patent information, such as controlling text generation with metadata (Lee, 2020), generating descriptions from drawings (Shukla et al., 2025), and producing specifications based on claims and drawings (Wang et al., 2024a). More recently, focus has shifted toward generating full patent documents from drafts (Wang et al., 2024b). For long-form patent generation scenarios, Knappich et al. (Knappich et al., 2025) pointed out that end-to-end supervised fine-tuning (SFT) tends to cause generation collapse and pathological repetition. To avoid such optimization instability, state-of-the-art methods generally adopt a train-free “divide-and-conquer” strategy: AutoPatent (Wang et al., 2024b)

uses a multi-agent framework to generate full documents from drafts, while Pap2Pat (Knappich et al., 2025) employs a chunk-based generation strategy. Differently, we investigate how to achieve stable, high-quality end-to-end paper-to-patent generation through RL training.

Alignment of Large Language Models. The paradigm for aligning large language models with human preferences has undergone significant changes. Early approaches primarily relied on reinforcement learning from human feedback (RLHF) (Stiennon et al., 2020; Bai et al., 2022), using proximal policy optimization (PPO) (Schulman et al., 2017) to maximize learned reward functions. To avoid the instability and complexity of online reinforcement learning, Direct Preference Optimization (DPO) (Rafailov et al., 2023) proposed a stable offline alternative that directly optimizes policies on preference data. This offline paradigm has since been improved through various techniques, including rejection sampling, listwise ranking, safety constraints, and theoretical generalization (Dong et al., 2023; Song et al., 2024; Liu et al., 2024; Dai et al., 2024; Azar et al., 2024). In recent years, research focus has shifted to iterative alignment strategies to address the limitations of static offline data. Such iterative paradigms have achieved notable successes in complex domains, such as enhancing multi-step reasoning abilities (Pang et al., 2024) and automating the full lifecycle of scientific discovery (Weng et al., 2025), whereas it has not been adopted in the paper-to-patent task.

3 Methodology

As illustrated in Figure 1, our training pipeline consists of two sequential stages: (1) supervised fine-tuning (SFT), which establishes foundational capabilities for factual adherence and patent-style writing; and (2) iterative preference optimization (Pang et al., 2024; Yuan et al., 2024), designed to align generated content with the factual consistency and linguistic style of authentic patents, while mitigating the pathological repetition issues.

3.1 Task Definition

Formally, given a dataset $\mathcal{D} = \{(x, y^*)\}$, where x denotes the source academic paper and y^* represents the corresponding ground-truth patent description, our goal is to learn a policy π_θ that generates patent description text y faithfully reflecting the technical facts in x , while adhering to the

language norms established by the United States Patent and Trademark Office (USPTO). We represent the end-to-end paper-to-patent generation process as follows:

$$[\mathcal{I}; x] \xrightarrow{\pi_\theta} y \quad (1)$$

where $[\mathcal{I}; x]$ denotes the concatenation of a carefully designed system prompt $\mathcal{I}$ (see Appendix A, Table 6) and the input paper x .

3.2 PatAlign

We first initialize our framework via standard supervised fine-tuning (SFT) to adapt a general-purpose large language model to the patent domain. Given the dataset $\mathcal{D} = \{(x, y^*)\}$, we minimize the negative log-likelihood of the reference patent y^* conditioned on the paper x and instruction $\mathcal{I}$, thereby obtaining an initial policy π_{SFT} .

However, as noted in prior work (Knappich et al., 2025), SFT on long-form patent texts tends to induce pathological repetition. This issue arises because the model struggles to distinguish between the necessary structural repetitions inherent to patent descriptions and degenerate pathological repetition.

While Direct Preference Optimization (DPO) (Rafailov et al., 2023) provides a mechanism to alleviate pathological repetition by introducing pathologically repetitive patent texts as negative samples, we argue that the conventional static (non-iterative) alignment paradigm is fundamentally limited for patent generation. Because negative samples typically consist of pathologically repetitive texts generated by early SFT checkpoints, the optimization objective persistently focuses on eliminating repetitive features, thereby suppressing the structural repetitions necessary to maintain patent rigor. Ultimately, this over-correction leads to the degradation of the patent language style.

To address this, we propose PATALIGN, a dynamic training framework based on iterative preference optimization. Unlike static DPO, we construct a curriculum in which the distribution of negative samples evolves synchronously with the policy. By utilizing the patent description texts generated by the current policy model based on the source papers as negative samples in each iteration, the optimization target always focuses on the *most salient* failure modes of the current policy—initially suppressing pathological repetition and gradually refining subtle style deviations.

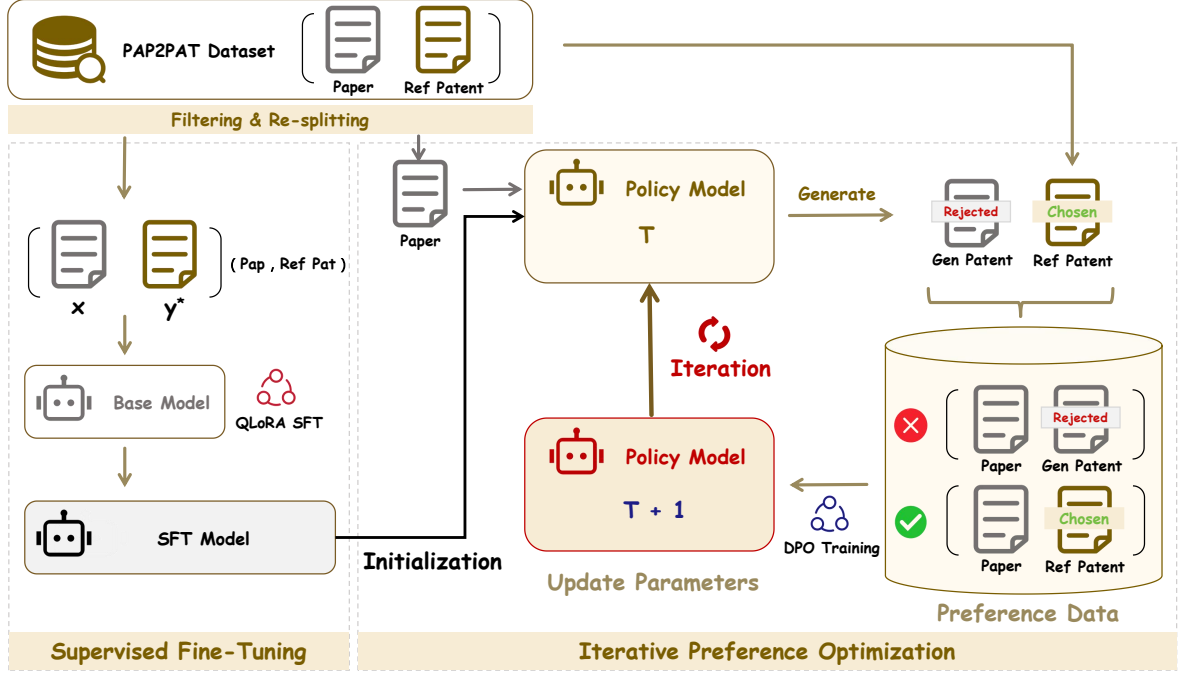


Figure 1: The overall framework of PatAlign consists of two stages: **Supervised Fine-Tuning (SFT)**, where the base model is fine-tuned on the filtered Pap2Pat dataset to establish foundational patent writing capabilities. **Iterative Preference Optimization**, where the model serves as the policy to generate patent description drafts.

Dynamic Preference Construction At iteration k (where $k = 1, \dots, K$), we construct a new preference dataset $\mathcal{D}_k$ using the current policy $\pi_{\theta_{k-1}}$. For each sample $(x, y^*) \in \mathcal{D}$, the ground-truth patent y^* serves as the preferred response y_w , and a rejected response y_l is sampled from the current policy: $\hat{y} \sim \pi_{\theta_{k-1}}(\cdot|x, \mathcal{I})$. This forms a preference triplet:

$$(x, y_w, y_l) = (x, y^*, \hat{y}) \quad (2)$$

This dynamic sampling is critical as $\hat{y}$ reflects the current model’s specific failure patterns (e.g., local repetition or style inconsistency), providing precise gradient signals for subsequent optimization.

Policy Optimization Starting from $\pi_{\theta_0} = \pi_{\text{SFT}}$, we perform K rounds of iterative updates. At each iteration, we optimize the policy to maximize the marginal probability of the preferred patent over the rejected sample, with the loss function defined as:

$$\mathcal{L}_{\text{DPO}}(\pi_{\theta_k}; \pi_{\text{ref}}) = -\mathbf{E}_{(x, y_w, y_l) \sim \mathcal{D}_k} \left[\log \sigma \left(\beta \log \frac{\pi_{\theta_k}(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta \log \frac{\pi_{\theta_k}(y_l|x)}{\pi_{\text{ref}}(y_l|x)} \right) \right] \quad (3)$$

where σ is the sigmoid function, and β controls the deviation penalty relative to the reference model.

Split	Raw (Original)	Filtered (Step 1)	Final (Step 2)
Train	1,000	741	910
Validation	242	172	100
Test	500	397 [†]	300
NC-Test	71	-	-
Total	1,813	1,310	1,310

Table 1: Statistics of dataset filtering and re-splitting.[†] indicates merging the NC-Test split into the filtered Test split.

A key component of our framework is the dynamic updating of the reference model: in the initial iteration, we set $\pi_{\text{ref}} = \pi_{\text{SFT}}$, and in subsequent iterations, π_{ref} is assigned to the policy from the previous iteration $\pi_{\theta_{k-1}}$. This rolling reference mechanism prevents catastrophic forgetting caused by excessive update magnitudes and avoids being overly constrained by the initial distribution, thereby ensuring stable style alignment training.

4 Implementation

We train PatAlign based on the OpenRLHF framework (Hu et al., 2025), using Qwen3-8B (Yang et al., 2025) as the backbone model.

4.1 Experimental Setup

Dataset. To enable end-to-end training, we construct the training, validation, and test sets for the paper-to-patent task based on the PAP2PAT dataset (Knappich et al., 2025), which contains paired academic papers and reference patents. To accommodate long-form generation and ensure data quality, we apply strict filtering and re-splitting criteria. Specifically, we remove samples whose combined paper–patent length exceeds 31,000 tokens to fit the context window of current large language models, and we also discard samples in which either the paper or the patent contains fewer than 1,000 tokens to avoid uninformative data. Since (Knappich et al., 2025) found pre-training contamination to have negligible impact on this task, we merge the non-contaminated test set (nc-test) from the original dataset into the test split for unified evaluation. In addition, to ensure that sufficient data are available for model training, we retain only 100 validation samples and 300 test samples, while assigning all remaining data to the training set. The final dataset therefore consists of 910 training samples, as detailed in Table 1. All results reported in this paper are evaluated on these 300 test samples.

Baselines. We compare the proposed method with three categories of baselines. (1) **Zero-shot Open-source Models:** To support long-form patent generation, we evaluate a set of open-source large language models that provide context windows of at least 32K tokens. These models primarily include the Qwen3 series (8B, 14B, 30B-A3B, and 32B) (Yang et al., 2025), as well as Llama-3.1-8B-Instruct (Grattafiori et al., 2024). All open-source models are prompted using an identical and carefully designed system prompt (see Table 6 in Appendix A), which emphasizes factual accuracy and patent-style writing, and directly generate patent descriptions from the input papers in a zero-shot setting. (2) **COPGEN** (Knappich et al., 2025): As a strong domain-specific baseline, we reproduce COPGEN by strictly following its official open-source implementation and reported hyperparameter configurations. This method generates patent text in a chunk-wise manner based on reference patent outlines and the source paper, and subsequently concatenates the generated segments into a complete document. (3) **Frontier Models:** We further evaluate two frontier models, GPT-5.5 and DeepSeek-V4-Pro (Xu et al., 2026). Given the substantial generation length involved in the

paper-to-patent task, we randomly sample 50 instances from the test set to control API evaluation costs. Given their support for long-context inputs, we evaluate both models under zero-shot, 1-shot, and 3-shot settings to investigate how the number of few-shot demonstrations affects paper-to-patent generation quality. Few-shot demonstrations are retrieved exclusively from the training set. For each test instance, we use BGE-M3 (Chen et al., 2024) to retrieve the most semantically similar paper from the training set and use its corresponding paper–patent pair as a demonstration. Both frontier models use the same retrieved demonstrations.

Evaluation Metrics. We follow the evaluation framework proposed in PAP2PAT (Knappich et al., 2025) to systematically assess generated patents from both content-level and language-level perspectives. (1) **Content-level metrics:** We use Coverage to measure the extent to which the generated patent covers the informational content of the reference patent. We further adopt Factuality metrics, including $Pap \rightarrow Gen$, $Ref \rightarrow Gen$, and $Ref+Pap \rightarrow Gen$. As our primary task is to perform factual style transfer grounded strictly in the source paper, we designate $Pap \rightarrow Gen$ as the principal indicator of factuality, whereas the other metrics are provided solely for supplementary reference. (2) **Language-level metrics:** We adopt the Style metric to quantify the alignment of the generated text with patent-specific writing conventions. In addition, to measure generation stability, we report the repetition rate (RR) and the proportion of windows with RR exceeding 80% ($RR_{>80}$). (3) **Paper-to-Patent composite metric:** To comprehensively evaluate the overall quality of the paper-to-patent transformation, we propose the Paper-to-Patent Score (P^2S). This composite metric is defined as the geometric mean of source-grounded factual consistency and language style:

$$P^2S = \sqrt{\text{Fac}_{Pap \rightarrow Gen} \cdot \text{Style}} \quad (4)$$

We adopt the geometric mean rather than the arithmetic mean to prevent compensation effects, where extremely high performance in one dimension masks severe deficiencies in another. This demands that a high-quality patent generation system must excel in both factuality and language style.

4.2 Implementation Details

Training Configurations. Training proceeds under a strict two-stage paradigm. *Supervised Fine-Tuning (SFT) Stage.* We utilize QLoRA (Dettmers

et al., 2023) for efficient fine-tuning, setting the LoRA rank to 64, learning rate to 1×10^{-4} , and global batch size to 32. Training is performed for 10 epochs on 4 NVIDIA A100-40G GPUs. The epoch-10 checkpoint is used as the SFT baseline reported in the main experiments. *Iterative Preference Optimization Stage.* We conduct 3 rounds of iterative DPO training. Notably, this stage is initialized from an intermediate SFT checkpoint (epoch 5) rather than the fully converged model to preserve policy plasticity and avoid distribution collapse caused by overfitting, thereby facilitating effective preference learning. For a controlled comparison, the *w/o RL* result reported in the ablation study corresponds to this same epoch-5 checkpoint. Each iteration runs for 1 epoch with a learning rate of 5×10^{-6} and KL penalty coefficient $\beta = 0.05$. Training is conducted on four NVIDIA RTX PRO 6000 Blackwell GPUs.

Inference Settings. Considering the strict factual consistency requirements of patent documents, we adopt a low sampling temperature of 0.3 during inference. To support the generation of full-length patent description, the maximum generation length is set to 30,000 tokens.

5 Results

Table 2 presents a detailed comparison of PATALIGN with the Skyline, zero-shot open-source models, and COPGEN. We establish a “Skyline” by evaluating the ground-truth reference patents. Because human-authored patents incorporate technical extensions beyond the source paper, this Skyline serves strictly as an upper bound for patent style rather than factuality, which is primarily measured against the source paper (*Pap*→*Gen*). Overall, PATALIGN achieves the highest composite P²S score of 82.7%, establishing new state-of-the-art results.

5.1 Main Results

Superiority in Factuality and Style. The central challenge of the Paper-to-Patent task is to simultaneously ensure faithfulness to the technical facts of the source paper and conformity with the specialized language style of patents. As shown in Table 2, PATALIGN demonstrates clear advantages on both dimensions, achieving the highest composite P²S score of 82.7% and establishing new state-of-the-art results. In terms of factuality (*Pap*→*Gen*), PATALIGN achieves a score of 78.7%, outperform-

ing the COPGEN baseline (53.2%) by 25.5 percentage points. With respect to language style, PATALIGN obtains a high score of 86.8%, substantially outperforming both COPGEN (65.9%) and the Qwen3-8B base model (63.7%).

Although COPGEN achieves a *Ref*→*Gen* score of 62.1% by benefiting from the inclusion of reference patent outlines during inference, this outline-guided strategy leads to deviations from the core technical facts of the input paper. Furthermore, although equipping COPGEN with a DPO-aligned model improves its Style score to 69.5%, its linguistic style remains fundamentally constrained compared to the end-to-end generation paradigm empowered by iterative preference optimization in PATALIGN. This limitation arises primarily because the chunk-based generation strategy fragments the document context, making it difficult to form a coherent and holistic patent writing style. In contrast, through iterative preference optimization, PatAlign remains faithful to the technical facts of the input paper while closely approximating the language style of real-world reference patents, which reach 95.7% under the Skyline setting.

Mitigating Pathological Repetition. We further analyze the role of PATALIGN in mitigating the issue of pathological repetition. As shown in Table 2, although the supervised fine-tuning (SFT) stage equips the model with foundational factual adherence and basic patent stylistic capabilities, yielding a P²S score of 80.5%, the model still suffers from severe pathological repetition during long-form generation, with the proportion of severely repetitive windows ($RR_{>80}$) reaching 32.2%. Furthermore, when adopting the chunk-based generation strategy of the COPGEN baseline, both Qwen3-8B SFT and DPO models exhibit even more severe pathological repetition issues, primarily because multiple inference passes significantly increase the frequency of triggering pathological repetition.

By employing iterative preference optimization, PATALIGN utilizes the model’s own pathologically repetitive outputs as dynamic negative samples, thereby providing precise gradient penalties against pathological repetition, which enables the model to escape the local optima induced by the likelihood maximization objective. Consequently, PATALIGN successfully reduces $RR_{>80}$ from 32.2% to 22.3%, corresponding to a decrease of 9.9 percentage points. Notably, this reduction is not achieved at the expense of content quality. Instead,

	Tokens	Coverage $\uparrow$	Factuality $\uparrow$			Style $\uparrow$	Repetitions		P ² S $\uparrow$
		<i>Gen</i> $\rightarrow$ <i>Ref</i>	<i>Pap</i> $\rightarrow$ <i>Gen</i>	<i>Ref</i> $\rightarrow$ <i>Gen</i>	<i>Ref</i> + <i>Pap</i> $\rightarrow$ <i>Gen</i>		RR	RR>80 $\downarrow$	
Skyline									
Reference Patent	11.5k	89.1 $\pm$ 5.83	47.8 $\pm$ 17.01	89.2 $\pm$ 5.30	89.4 $\pm$ 5.28	95.7	14.1	0.1	67.6
Zero-shot Open-source Models									
Llama-3.1-8B-Instruct	3.5k	36.6 $\pm$ 10.47	68.0 $\pm$ 15.05	49.9 $\pm$ 18.22	68.9 $\pm$ 15.13	71.4	19.9	0.8	69.7
Qwen3-8B	2.1k	38.0 $\pm$ 11.31	65.6 $\pm$ 11.17	48.8 $\pm$ 17.29	66.4 $\pm$ 11.12	63.7	7.7	0.0	64.6
Qwen3-14B	2.5k	39.6 $\pm$ 11.52	61.6 $\pm$ 11.14	46.2 $\pm$ 15.08	63.2 $\pm$ 11.14	66.9	7.7	0.0	64.2
Qwen3-30B-A3B	2.4k	39.3 $\pm$ 11.50	60.9 $\pm$ 12.22	46.4 $\pm$ 16.09	61.7 $\pm$ 11.94	66.8	6.5	0.0	63.8
Qwen3-32B	2.1k	38.8 $\pm$ 11.39	57.1 $\pm$ 11.30	43.2 $\pm$ 14.07	58.3 $\pm$ 11.14	63.7	6.6	0.0	60.3
COPGEN (Knappich et al., 2025)									
Llama-3.1-8B-Instruct	7.6k	40.9 $\pm$ 12.38	49.6 $\pm$ 13.52	60.9 $\pm$ 12.64	62.3 $\pm$ 12.83	70.6	25.0	3.0	59.2
Qwen3-8B	3.3k	40.3 $\pm$ 12.46	53.2 $\pm$ 14.13	62.1 $\pm$ 12.18	64.4 $\pm$ 11.59	65.9	7.4	0.0	59.2
Qwen3-8B SFT	29.3k	39.3 $\pm$ 11.41	45.9 $\pm$ 15.95	51.6 $\pm$ 17.16	53.0 $\pm$ 16.93	67.9	74.3	56.9	55.8
Qwen3-8B DPO	27.3k	39.6 $\pm$ 12.30	47.8 $\pm$ 17.88	54.2 $\pm$ 18.78	55.7 $\pm$ 18.65	69.5	73.4	56.8	57.6
Qwen3-14B	5.5k	42.3 $\pm$ 12.52	50.1 $\pm$ 12.22	56.8 $\pm$ 11.11	58.8 $\pm$ 11.23	67.7	7.9	0.0	58.2
Qwen3-30B-A3B	4.8k	42.1 $\pm$ 11.58	51.6 $\pm$ 13.69	59.2 $\pm$ 11.25	61.5 $\pm$ 11.18	67.1	6.8	0.0	58.8
Qwen3-32B	6.5k	42.9 $\pm$ 12.86	48.6 $\pm$ 13.29	56.4 $\pm$ 11.81	58.1 $\pm$ 11.83	70.0	7.0	0.4	58.3
PatAlign (ours)									
Qwen3-8B SFT	15.4k	43.2 $\pm$ 15.52	74.8 $\pm$ 20.62	47.4 $\pm$ 21.50	76.1 $\pm$ 19.74	86.6	41.0	32.2	80.5
Qwen3-8B PatAlign	12.7k	42.0 $\pm$ 14.64	78.7 $\pm$ 17.03	48.1 $\pm$ 20.84	79.2 $\pm$ 17.47	86.8	31.9	22.3	82.7

Table 2: Main results comparing Skyline, Zero-shot Open-source models, COPGEN, and our proposed methods. The evaluation covers Coverage and Length (Tokens, Coverage), Factual Faithfulness with respect to the source paper and reference patent (*Pap* $\rightarrow$ *Gen*, *Ref* $\rightarrow$ *Gen*, *Ref*+*Pap* $\rightarrow$ *Gen*), Patent-Style Alignment (Style), and Long-Form Generation Stability (RR, RR>80). P²S denotes the overall Paper-to-Patent Score. Tokens and RR achieve optimal performance when closest to the reference patent. Best results are highlighted in **bold** with orange background.

PATALIGN further improves the *Pap* $\rightarrow$ *Gen* factuality score from 74.8% to 78.7%. This demonstrates that our method not only effectively suppresses pathological repetition but also actively guides the model to concentrate probability mass on token sequences that are more faithful to the technical facts of the source paper and closer to authentic patent language styles.

Small Models Beat Larger-size Baselines. Our results highlight the importance of domain-specific alignment, showing that a carefully aligned small model can substantially outperform general-purpose large language models with significantly more parameters. As shown in Table 2, PATALIGN, with 8B parameters, consistently outperforms the Qwen3-32B base model—which has approximately four times the parameter count—across all core metrics.

To eliminate the impact of different generation strategies, we compare both models under an identical end-to-end generation setting. Specifically, the general-purpose Qwen3-32B achieves a P²S score of 60.3%, a *Pap* $\rightarrow$ *Gen* factuality score of 57.1%, and a language Style score of 63.7%. In

contrast, PATALIGN yields absolute gains of 22.4, 21.6, and 23.1 percentage points in P²S, *Pap* $\rightarrow$ *Gen* factuality, and Style, respectively. These results provide strong evidence that, for the highly specialized paper-to-patent task, simply scaling up the parameter count of general-purpose models fails to yield effective performance improvements; instead, targeted iterative alignment training on a smaller model is more effective. Consequently, PATALIGN offers an efficient pathway for deploying high-quality paper-to-patent generation systems under constrained computational resources.

5.2 Comparison with Frontier Models

As shown in Table 4, few-shot prompting improves the generation performance of both frontier models, while the marginal gains diminish as more demonstrations are provided.

Despite these gains from few-shot prompting, PATALIGN retains a clear advantage. Compared with the strongest frontier-model setting, GPT-5.5 (3-shot), PATALIGN achieves gains of 6.2, 7.3, and 6.8 percentage points in source-grounded factuality (Fac_{Pap} $\rightarrow$ *Gen*), Style, and P²S, respectively.

	Tokens	Coverage $\uparrow$	Factuality $\uparrow$			Style $\uparrow$	Repetitions		P ² S $\uparrow$
		<i>Gen</i> $\rightarrow$ <i>Ref</i>	<i>Pap</i> $\rightarrow$ <i>Gen</i>	<i>Ref</i> $\rightarrow$ <i>Gen</i>	<i>Ref</i> + <i>Pap</i> $\rightarrow$ <i>Gen</i>		RR	RR>80 $\downarrow$	
PatAlign	12.7k	42.0 $\pm$ 14.64	78.7 $\pm$ 17.03	48.1 $\pm$ 20.84	79.2 $\pm$ 17.47	86.8	31.9	22.3	82.7
w/o Iterative	12.5k	42.5 $\pm$ 15.32	78.6 $\pm$ 16.25	48.7 $\pm$ 20.76	79.0 $\pm$ 16.80	85.7	33.6	21.9	82.1
w/o RL	15.4k	42.5 $\pm$ 15.68	74.3 $\pm$ 20.75	47.3 $\pm$ 20.71	75.5 $\pm$ 19.87	87.6	42.7	32.8	80.7
Base Model	2.1k	38.0 $\pm$ 11.31	65.6 $\pm$ 11.17	48.8 $\pm$ 17.29	66.4 $\pm$ 11.12	63.7	7.7	0.0	64.6

Table 3: Ablation study: we examine the contribution of each component. Specifically, **w/o Iterative** removes the iterative optimization procedure, reducing the method to standard DPO, while **w/o RL** further removes preference-based alignment, effectively reducing the model to supervised fine-tuning only.

Model	Setting	Fac _{Pap→Gen}	Style	P ² S
DeepSeek-V4-Pro	0-shot	53.3	71.8	61.9
	1-shot	54.8	73.6	63.5
	3-shot	55.6	73.2	63.8
GPT-5.5	0-shot	64.2	66.4	65.3
	1-shot	73.5	67.2	70.3
	3-shot	73.8	67.6	70.6
PatAlign	0-shot	80.0	74.9	77.4

Table 4: Comparison with frontier models on the same 50 randomly sampled test instances. “Setting” denotes the number of demonstrations provided during inference.

These results demonstrate that, for paper-to-patent generation requiring strict factual consistency and patent-specific language style, targeted alignment training enables an 8B model to outperform few-shot prompting with frontier models.

5.3 Ablation Study

Table 3 systematically presents the impact of different training strategies. Although Direct Preference Optimization (w/o Iterative) effectively mitigates the pathological repetition introduced by supervised fine-tuning (w/o RL), it brings noticeable style degradation. As shown in Table 3, applying DPO reduces the Style score from 87.6% to 85.7%. This degradation occurs because static negative samples are drawn from pathologically repetitive patent texts generated by the SFT checkpoint. Consequently, the optimization target persistently focuses on eliminating repetition. As training progresses and the model’s generation distribution learns to avoid pathological repetition, this unchanging optimization target begins to penalize the necessary structural repetitions in patent texts, ultimately leading to the degradation of the patent language style. In contrast, by dynamically constructing negative samples, PATALIGN ensures that

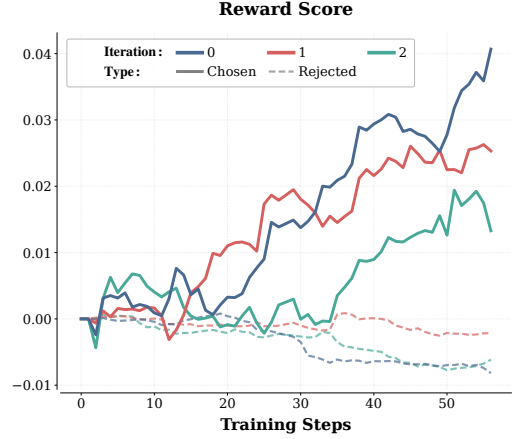


Figure 2: Reward trajectories of chosen and rejected samples across three iterative cycles of PATALIGN.

each iteration targets the most prominent generation issues of the current model for optimization. This evolving optimization target effectively alleviates the style degradation caused by DPO, improving the Style score to 86.8%.

5.4 Iterative Reward Dynamics

To investigate why the iterative optimization framework outperforms its non-iterative counterpart, we analyze the evolution of rewards during training. Figure 2 illustrates the reward trajectories of the chosen and rejected samples across three iterative cycles of PATALIGN. We observe a pronounced Cyclic Reset–Divergence pattern: at the beginning of each iteration, the reward margin between the chosen and rejected samples exhibits a clear convergence. This phenomenon provides strong evidence that the previous iteration has successfully pulled the policy model’s generation distribution closer to the reference patent distribution.

However, as the training process of each iteration progresses, the reward margin rapidly widens again. This trend demonstrates that, despite the increased discrimination difficulty induced by these increas-

Method	Fact.	Style	Disclosure	Overall	P ² S
COPGEN (8B)	5.6	6.4	6.6	6.2	57.6
GPT-5.5	7.5	7.1	8.1	7.6	65.3
PatAlign (8B)	8.3	7.9	8.4	8.2	77.9

Table 5: Human evaluation on 20 test instances randomly sampled from the 50-instance frontier-model evaluation subset. Disclosure Sufficiency measures whether the generated patent adequately describes the core invention, key components, data flows, technical mechanisms, and implementable embodiments. Overall denotes the average of Factuality, Style, and Disclosure Sufficiency, while P²S is computed on the same 20 instances.

ingly higher-quality negative samples, PATALIGN still enables the model to gradually achieve an effective distinction between positive and negative samples during each training cycle. Through this curriculum-learning manner, the model continuously uncovers finer-grained linguistic style differences between its own generated negative samples and the authentic reference patents, thereby iteratively aligning with the reference patent distribution.

5.5 Human Evaluation

We further conduct a human evaluation. Each sample is assessed by the same evaluator with experience in patent drafting on a 1–10 scale to maintain a consistent evaluation standard across models. Model identities are anonymized and the presentation order of outputs is randomly shuffled to reduce potential bias from model identity and ordering.

As shown in Table 5, PATALIGN achieves the highest human scores across Factuality, Style, and Disclosure Sufficiency, resulting in an Overall score of 8.2, compared with 7.6 for GPT-5.5 and 6.2 for COPGEN. Meanwhile, the ranking of the patents generated by the three methods according to P²S is fully consistent with that based on the human Overall scores, indicating strong agreement between P²S and expert judgments of paper-to-patent generation quality and supporting its use as an automated aggregate evaluation metric for this task.

5.6 Training Efficiency Analysis

As detailed in Table 7 of Appendix B, although PATALIGN introduces multiple rounds of iterative preference optimization, it effectively alleviates the style degradation issue of static DPO at an accept-

able computational cost. Compared to the static DPO pipeline, it incurs no higher peak GPU memory overhead, and the total training time increases by only 3.8 hours.

5.7 Case Study

Pathological Repetition. As illustrated in Figure 3 in Appendix C, the SFT baseline model suffers from severe pathological repetition during patent generation (e.g., repeatedly generating “*The invention also relates to ... The invention also relates to ...*”). In contrast, PATALIGN effectively mitigates this pathological repetition phenomenon and generates logically coherent patent texts.

Factuality and Language Style. To further evaluate generation quality, we present a comparative case study between PATALIGN (Qwen3-8B) and the baseline COPGEN (Qwen3-32B) in Appendix C. As illustrated in Figure 4, COPGEN suffers from factual hallucinations (e.g., fabricating non-existent entities like “A β MRF”), and its generated text still strongly exhibits an academic paper writing style. Conversely, PATALIGN not only maintains factual consistency with the source paper but also generates texts that are much closer to the authentic patent language style.

6 Conclusion

In this paper, we proposed PATALIGN, an iterative direct preference optimization framework for the paper-to-patent generation task. PATALIGN effectively resolves the pathological repetition in supervised fine-tuning and the style degradation in static preference optimization. Additionally, we introduced P²S, a comprehensive metric for evaluating overall patent generation quality. Experimental results demonstrate that PATALIGN establishes a new state-of-the-art by achieving both high factual consistency and an authentic patent writing style. We believe this work not only improves the drafting efficiency of patent attorneys but also provides valuable insights for aligning long-form text generation models under complex stylistic constraints.

Limitations

We acknowledge several limitations in our study, which concurrently highlight promising directions for future research. First, our primary objective is to provide patent drafting assistance rather than directly generating applications with final legal enforceability. Future work could integrate this frame-

work into a human-in-the-loop workflow, combining the model’s efficient generation capabilities with expert review and revision to ultimately produce complete and legally robust application documents. Second, constrained by the context window limits of current open-source models, we filtered out document pairs exceeding 31k tokens. As open-source foundation models continue to advance in context length, our approach can be naturally extended to ultra-long patent generation scenarios. Furthermore, because currently available public paper-patent paired datasets are restricted to USPTO standards, our training and evaluation are conducted solely based on this jurisdiction. Fundamentally, PATALIGN serves as a general alignment framework; should data from other countries or regions become publicly available, we can readily explore its adaptation to the patent standards of other jurisdictions. Finally, limited by computational resources, we have not yet explored the alignment effectiveness on larger-scale foundation models. We leave this extended investigation into scaling behaviors for future work.

Ethics Statement

The data utilized in this study are derived from the publicly available PAP2PAT dataset, comprising paired academic papers and patent documents from open-access sources. The primary objective of our proposed framework, PATALIGN, is to assist patent professionals by automating the conversion of scientific disclosures into initial patent description drafts, thereby enhancing the efficiency of intellectual property protection. We explicitly condemn the misuse of this technology for malicious activities, such as generating fraudulent applications, spamming, or facilitating patent trolling, which could impose undue burdens on intellectual property offices. We emphasize that the outputs generated by our model serve only as auxiliary references and do not constitute final legal documents. Despite improvements in factual consistency and stylistic alignment, the generated descriptions require rigorous verification and modification by qualified patent agents to ensure they meet strict legal standards and technical accuracy.

Acknowledgments

We thank the reviewers for their constructive comments and suggestions, which helped improve this work. This work was supported by the High-level

Special Funds (No.G03050K007) of Southern University of Science and Technology.

References

- Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. 2024. A general theoretical paradigm to understand learning from human preferences. In *International Conference on Artificial Intelligence and Statistics*, pages 4447–4455. PMLR.
- Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, and 1 others. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.
- Silvia Casola and Alberto Lavelli. 2022. Summarization, simplification, and generation: The case of patents. *Expert Systems with Applications*, 205:117627.
- Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, and 1 others. 2024. A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, 15(3):1–45.
- Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*, 4(5).
- Joan Codina-Filbà, Nadjat Bouayad-Agha, Alicia Burga, Gerard Casamayor, Simon Mille, Andreas Müller, Horacio Saggion, and Leo Wanner. 2017. Using genre-specific features for patent summaries. *Information Processing & Management*, 53(1):151–174.
- Juntao Dai, Xuehai Pan, Ruiyang Sun, Jiaming Ji, Xinbo Xu, Mickel Liu, Yizhou Wang, and Yaodong Yang. 2024. Safe rlhf: Safe reinforcement learning from human feedback. In *International Conference on Learning Representations*, volume 2024, pages 50750–50777.
- Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36:10088–10115.
- Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. 2023. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten,

- Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Jian Hu, Xibin Wu, Wei Shen, Jason Klein Liu, Weixun Wang, Songlin Jiang, Haoran Wang, Hao Chen, Bin Chen, Wenkai Fang, and 1 others. 2025. Openrlhf: A ray-based easy-to-use, scalable and high-performance rlhf framework. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 656–666.
- Lekang Jiang, Chengzu Li, and Stephan Goetz. 2025a. Enriching patent claim generation with european patent dataset. *arXiv preprint arXiv:2505.12568*.
- Lekang Jiang, Caiqi Zhang, Pascal A Scherz, and Stefan Goetz. 2025b. Can large language models generate high-quality patent claims? In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1272–1287.
- Sunhye Kim and Byungun Yoon. 2022. Multi-document summarization for patent documents based on generative adversarial network. *Expert Systems with Applications*, 207:117983.
- Valentin Knappich, Anna Hätt, Simon Razniewski, and Annemarie Friedrich. 2025. Pap2pat: benchmarking outline-guided long-text patent generation with patent-paper pairs. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 9524–9554.
- Jieh-Sheng Lee. 2020. Controlling patent text generation by structural metadata. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 3241–3244.
- Jieh-Sheng Lee. 2025. Instructpatentgpt: training patent language models to follow instructions with human feedback. *Artificial Intelligence and Law*, 33(3):739–782.
- Jieh-Sheng Lee and Jieh Hsiang. 2020. Patent claim generation by fine-tuning openai gpt-2. *World Patent Information*, 62:101983.
- Junhua LI, Qian YUAN, Xiang YAN, and Changhong LV. 2025. Research on automated generation and evaluation of patent claims based on gpt-4. *Scientific Information Research*, 7(1):9.
- Tianqi Liu, Yao Zhao, Rishabh Joshi, Misha Khalman, Mohammad Saleh, Peter Liu, and Jialu Liu. 2024. Statistical rejection sampling improves preference optimization. In *International conference on learning representations*, volume 2024, pages 54605–54624.
- Fiona Murray and Scott Stern. 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization*, 63(4):648–687.
- Richard Yuanzhe Pang, Weizhe Yuan, He He, Kyunghyun Cho, Sainbayar Sukhbaatar, and Jason Weston. 2024. Iterative reasoning preference optimization. *Advances in Neural Information Processing Systems*, 37:116617–116637.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.
- Runtao Ren, Jian Ma, and Jianxi Luo. 2025. Large language model for patent concept generation. *Advanced Engineering Informatics*, 65:103301.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Shreya Shukla, Nakul Sharma, Manish Gupta, and Anand Mishra. 2025. Patentlm: large multimodal model for generating descriptions for patent figures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 20488–20496.
- Feifan Song, Bowen Yu, Minghao Li, Haiyang Yu, Fei Huang, Yongbin Li, and Houfeng Wang. 2024. Preference ranking optimization for human alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18990–18998.
- Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in neural information processing systems*, 33:3008–3021.
- Amy JC Trappey, Charles V Trappey, Jheng-Long Wu, and Jack WC Wang. 2020. Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Advanced Engineering Informatics*, 43:101027.
- Tony Trippe and Jieh-Sheng Jason Lee. 2025. Special issue on applications of generative ai and large language models in the patent domain.
- Yuen-Hsien Tseng, Chi-Jen Lin, and Yu-I Lin. 2007. Text mining techniques for patent analysis. *Information processing & management*, 43(5):1216–1247.
- Juanyan Wang, Sai Krishna Reddy Mudhiganti, and Manali Sharma. 2024a. Patentformer: a novel method to automate the generation of patent applications. In *Proceedings of the 2024 conference on empirical methods in natural language processing: industry track*, pages 1361–1380.
- Qiyao Wang, Shiwen Ni, Huaren Liu, Shule Lu, Guhong Chen, Xi Feng, Chi Wei, Qiang Qu, Hamid Alinejad-Rokny, Yuan Lin, and 1 others. 2024b. Autopatent: a multi-agent framework for automatic patent generation. *arXiv preprint arXiv:2412.09796*.

Yixuan Weng, Minjun Zhu, Guangsheng Bao, Hongbo Zhang, Jindong Wang, Yue Zhang, and Linyi Yang. 2025. Cycleresearcher: Improving automated research via automated review. In *International Conference on Learning Representations*, volume 2025, pages 3669–3709.

Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chencheng Ling, and 1 others. 2026. Deepseek-v4: Towards highly efficient million-token context intelligence. *arXiv preprint arXiv:2606.19348*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Naoya Yoshikawa and Ralf Krestel. 2024. Improving patent classification using ai-generated summaries. In *5th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech2024), Workshop At SIGIR*, volume 3775, pages 8–17. CEUR-WS. org.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. 2024. Self-rewarding language models. *arXiv preprint arXiv:2401.10020*.

Longhui Zhang, Lei Li, and Tao Li. 2015. Patent mining: a survey. *ACM Sigkdd Explorations Newsletter*, 16(2):1–19.

A System Prompts for Evaluation

Table 6 presents the complete system prompt and the associated output format constraints used during inference for the Base Model, SFT, DPO, and PatAlign methods.

B Training Efficiency Analysis

Table 7 details the computational resource consumption and time overhead of different training configurations. Although PATALIGN introduces multiple rounds of iterative preference optimization, its peak VRAM remains at 4×94.97 GB, incurring no higher peak GPU memory overhead compared to the static DPO pipeline. The total time encompasses both the policy training time and the inference time required to generate negative samples.

C Case Study Visualization

In this section, we provide the full visualization of the case studies discussed in the main text. We first analyze the pathological repetition issue in the SFT baseline, followed by a detailed comparison

of factuality and language style between COPGEN and PATALIGN.

C.1 Pathological Repetition Analysis

Figure 3 compares the generation outputs of the SFT baseline and our PATALIGN model given the same input scientific paper. The SFT model collapses into an infinite loop (highlighted in red) due to shortcut learning on patent redundancy. Conversely, PATALIGN avoids this degeneration, producing a coherent, logically progressive description that accurately reflects the source paper’s contributions.

C.2 Detailed Case Study on Factuality and Language Style

To qualitatively compare generation quality, Figure 4 presents a representative case study on a yeast reporter system for A β 42 oligomerization. We compare our proposed PATALIGN (Qwen3-8B) against the strong baseline COPGEN (Qwen3-32B) from two critical perspectives: factual consistency with the source paper and authentic reference patent language style alignment.

Factual Consistency with the Source Paper.

The source paper explicitly states that the proposed system is based on “a fusion of A β 42 to a reporter protein” in yeast. PATALIGN faithfully preserves this core scientific fact by generating: “The system includes a fusion protein comprising A β 42 and a reporter protein...”. This description remains factually consistent with the source paper while reformulating the content into an expression that conforms to the patent language style. In contrast, COPGEN introduces a severe factual hallucination by generating: “The model employs an A β 42-mammalian reporter fusion protein (A β MRF)...”. Neither the term “mammalian reporter fusion protein” nor the abbreviation “A β MRF” appears in the source paper. This demonstrates that COPGEN may introduce unsupported fabricated content during generation, thereby reducing factual consistency.

Patent-style Language Alignment. Beyond factual consistency, patent drafting requires a specialized patent language style that differs significantly from academic writing. The source paper uses standard academic discourse, such as “While such approaches may reduce plaque formation...” and “In contrast, the new model system allows...”. COPGEN largely preserves these academic rhetorical patterns in its generated text (e.g., “To address these

System Prompt Configuration
Role You are a senior patent attorney and technical expert specializing in drafting U.S. patent specifications. Task Transform the provided research paper content into a complete and fully developed patent specification. Guidelines Expansion and Elaboration: Build upon the core technical concepts presented in the paper and reasonably expand and elaborate them to produce a comprehensive patent specification. Describe implementation details thoroughly to meet the typical depth and length expected in U.S. patent documents. Consistency With Source Material: Ensure that all generated content remains consistent with the core technical facts, data, and disclosures contained in the input research paper. Patent-Style Writing: Use a professional, objective, and legally rigorous U.S. patent drafting style. Employ appropriate terminology, formal tone, and structured narrative consistent with standard patent specifications.
Output Format Constraint (Appended to User Prompt)
Output Format Requirements Content Constraints: Do not generate a “Claims” section and do not generate an Abstract. Focus solely on the specification description. Subsections Flexibility: You may add subsections (### ...) under <i>any</i> top-level or secondary heading whenever needed to improve structure or clarity. - You must output the content in Markdown format to maintain structural clarity. <i>The following rules and examples are provided as general reference only, not as strict required ordering:</i> Heading Hierarchy Reference - Use # for the Patent Title. - Use ## for Main Sections. - Use ### for Subsections under any section as needed. Example Section Headings (Reference Only) <pre># [Insert Patent Title Here] ## FIELD OF THE INVENTION ## BACKGROUND OF THE INVENTION ## SUMMARY OF THE INVENTION ## DETAILED DESCRIPTION OF THE INVENTION (You may introduce subsections—for example, ### System Architecture, ### Method Flow, or any other appropriate breakdown—under any section.) ### END OF PATENT ###</pre>

Table 6: Complete system prompt and output format constraints used for evaluating all models. The prompt specifies the model’s assumed role and enforces strict formatting and content guidelines during generation.

limitations...” and “*In contrast, the new model system allows...*”). Such expressions resemble the academic paper style rather than patent disclosure language, indicating that it struggles to generate text that conforms to the patent language style. In comparison, PATALIGN can transform the source paper facts into patent-style expressions, including “*The present invention provides...*”, “*The system comprises...*”, and “*The method comprises...*”. These declarative phrases demonstrate stronger alignment with patent drafting conventions.

Method	Peak VRAM	Train Time	Neg Gen Time	Total Time	Avg Time / Sample
PATALIGN	4 × 94.97 GB	14.9h	7.0h	21.9h	12.5s
w/o Iterative	4 × 94.97 GB	14.6h	3.5h	18.1h	12.5s
w/o RL	4 × 38.7 GB	11.9h	–	11.9h	–

Table 7: Training efficiency analysis. **w/o Iterative** refers to the standard sequential training pipeline of SFT followed by static DPO. “Neg Gen Time” denotes the total inference time required to generate negative samples, while “Avg Time/Sample” represents the average generation time per sample during this phase.

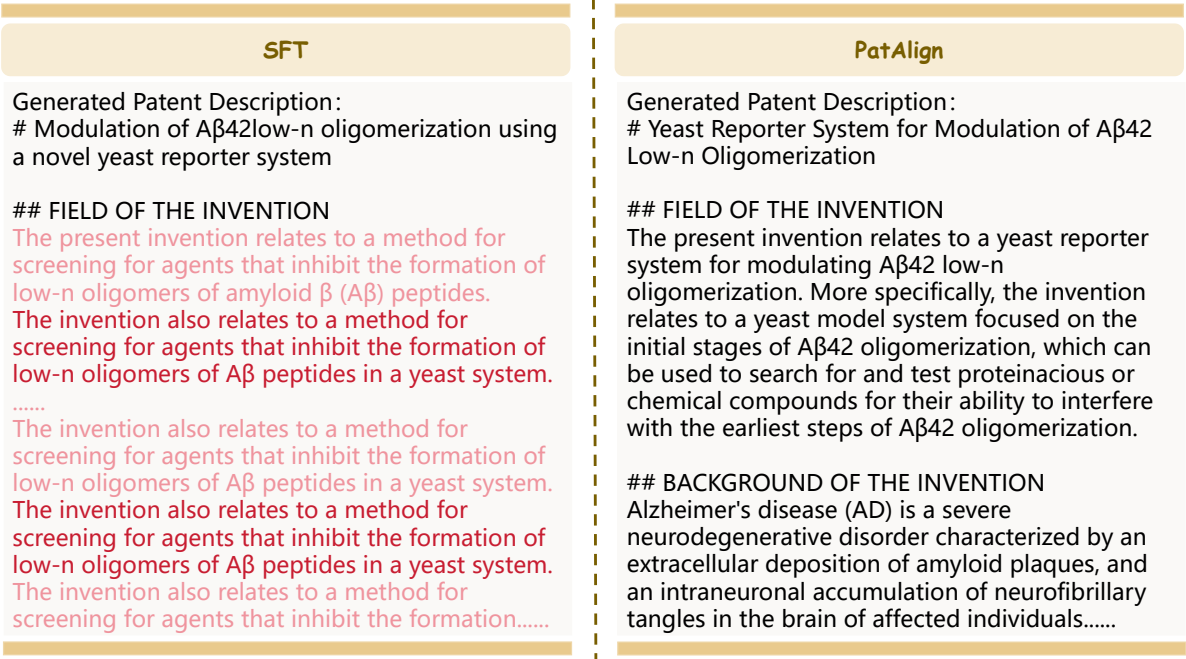


Figure 3: Case study comparing SFT and PATALIGN. The SFT model collapses into an infinite loop due to shortcut learning on patent redundancy. Conversely, PATALIGN avoids this degeneration, generating a logically coherent patent text.

Paper	COPGEN (Qwen3-32B)	PatAlign (Qwen3-8B)
Source Paper Evidence:	Baseline Generation:	Our Generation:
...We show that the activity of a fusion of Aβ42 to a reporter protein is compromised in yeast by the formation of SDS-stable low-n oligomers... While such approaches may reduce plaque formation, they often result in the accumulation of toxic oligomers, which may exacerbate disease progression. In contrast, the new model system allows for the investigation of compounds that interfere with the initial stages of Aβ42 oligomerization...	✗ Factual Hallucination ...The model employs an Aβ42-mammalian reporter fusion protein (AβMRF), which enables the detection of oligomerization through functional assays... To address these limitations, a novel yeast-based model system has been developed... In contrast, the new model system allows for the investigation of compounds... ✗ Academic Paper Style	...The system includes a fusion protein comprising Aβ42 and a reporter protein... ✓ Factual Consistency The present invention provides a yeast reporter system for modulating Aβ42 low-n oligomerization. The system comprises a yeast model system focused on the initial stages of Aβ42 oligomerization...The method comprises growing the yeast in a medium containing a compound... ✓ Patent-style Language
Academic Paper Style		

Figure 4: A comparative case study between COPGEN (Qwen3-32B) and PATALIGN (Qwen3-8B) evaluating factual consistency and language style.