RoSRL: Adaptive Rule-of-Sum Reinforcement Learning for Efficient and Reliable Summarization

Thu Phuong Tran Thi[1,2], Vinh Nguyen Van[2], Thai Nguyen Phuong[2], Quang Vu Ngoc[3], Khoa Nguyen Dang[4]

[1]Hanoi Metropolitan University, Hanoi, Vietnam

[2]VNU University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

[3]FPT IS Company Limited [4]University of Rochester

tttphuong2@daihocthudo.edu.vn vinhnv@vnu.edu.vn thainp@vnu.edu.vn

quang.vn@outlook.com knguy42@u.rochester.edu

Abstract

Hallucination remains a critical challenge for summarization, especially in domains such as law and large-scale news where factual accuracy is paramount. While reinforcement learning from human feedback (RLHF) and direct preference optimization (DPO) can reduce hallucination, they require costly preference annotations, limiting applicability in low-resource preference-label settings. Extractive methods inherently avoid unsupported content but often rely on static heuristics, constraining adaptability and coherence. We present RoSRL (Rule-of-Sum Reinforcement Learning), an efficient and reliable label-free extractive framework that combines interpretable, feature-based sentence scoring with reinforcement learning. RoSRL incorporates Lean-Proof Lite, a lightweight validator ensuring numerical and entity-level consistency, and adapts feature weights via multi-dimensional rewards—faithfulness, coverage, coherence, brevity, and section balance—optimized with Proximal Policy Optimization (PPO). Experiments on Vietnamese and English news and U.S. legal texts show consistent gains under ROUGE/BERTScore and other model-based summary evaluation methods, corroborated by human judgments. These results highlight RoSRL's efficiency, scalability, and reliability as a resource-conscious alternative to heuristic baselines and preference-optimized models, and underscore its potential as a foundation for future abstractive extensions.

1 Introduction

Automatic text summarization condenses large volumes of text into concise summaries that preserve essential information, enabling efficient information access in domains such as news, scientific publications, and legal documents. Existing approaches are typically categorized as extractive, which assemble summaries by selecting salient source sentences, and abstractive, which generate paraphrased content. Abstractive methods, powered by large neural language models, produce highly fluent summaries but are prone to factual errors. Extractive methods maintain stronger grounding in the source text, thus reducing the risk of hallucination, but often sacrifice coherence.

Hallucination remains a critical challenge for real-world deployment, especially in domains where factual reliability is paramount, such as law and high-volume news reporting. Neural abstractive summarizers frequently generate unsupported statements, making them unsuitable in contexts where factual errors carry legal or societal consequences (Maynez et al., 2020; Ji et al., 2023). Recent approaches such as RLHF (Stiennon et al., 2020; Ouyang et al., 2022) and direct preference optimization (DPO) (Rafailov et al., 2023) have shown promise in reducing hallucination, but both require large-scale annotated preference datasets, which are expensive to produce and scarce in low-resource languages like Vietnamese.

Extractive summarization provides a lightweight alternative by directly composing summaries from source sentences, inherently lowering the hallucination risk. However, sentence misplacement or incoherent selection can still distort meaning if taken out of context. Classic extractive methods such as Maximal Marginal Relevance (MMR) (Carbonell and Goldstein, 1998), LexRank (Erkan and Radev, 2004), and TextRank (Mihalcea and Tarau, 2004) rely on manually tuned heuristics (e.g., sentence centrality, redundancy reduction) that limit adaptability to domain-specific needs for factual consistency and coherence. Neural extractive methods such as BERTSum (Liu and Lapata, 2019) leverage pretrained language models to improve performance but remain supervised and require labeled summaries, limiting scalability and cross-domain applicability. More recent reference-free approaches like OTExtSum (Tang et al., 2022) optimize semantic coverage but do not explicitly

ensure factual reliability, structural balance, or multilingual generalization (Min et al., 2025; Adams et al., 2023).

Reinforcement learning (RL) offers a promising pathway to overcome these limitations. Models such as the Reinforced Neural Extractive Summarizer (RNES) (Wu and Hu, 2018) and BanditSum (Dong et al., 2018) demonstrate that optimizing informativeness and coherence via bandit or policy-gradient methods can improve extractive summarization without gold-standard supervision. However, existing RL-based extractive methods do not explicitly address multi-dimensional quality objectives—such as factual consistency, coverage, and section balance—and show limited generalization to multilingual and low-resource contexts.

To address these gaps, we propose RoSRL: Adaptive Rule-of-Sum Reinforcement Learning for Efficient and Reliable Summarization, a label-free extractive framework that unifies interpretable feature-based sentence scoring with policy optimization. RoSRL is designed to be efficient in computation and reliable in factuality. It integrates two components: (*i*) Lean-Proof Lite, a lightweight factuality validator that enforces numerical, entity, and semantic consistency; and (*ii*) ADAPT_RULE, a PPO-based adaptive mechanism that dynamically refines feature weights under multi-dimensional rewards covering faithfulness, coverage, coherence, brevity, and section balance. By removing the need for labeled summaries and optimizing directly for complementary quality dimensions, RoSRL attains competitive accuracy with modest resources across languages and domains.

Our contributions are as follows:

- We introduce RoSRL, an adaptive rule-of-sum reinforcement learning framework for extractive summarization that combines interpretable feature scoring with dynamic weight refinement and Lean-Proof Lite for numerical and entity-level consistency.
- We design multi-dimensional, reference-free reward signals that drive PPO-based adaptation, enabling balanced optimization across factuality, coverage, coherence, brevity, and section representation without reliance on costly preference annotations.
- Through extensive experiments on Vietnamese news (VNExpress), CNN/DailyMail, and U.S. legal texts (BillSum), we show that the proposed reward design and adaptive weighting yield consistent improvements over strong heuristic and unsupervised baselines in automatic metrics, model-based summary evaluation, and human judgments.
- We demonstrate that RoSRL operates efficiently with modest GPU resources, underscoring its practicality for multilingual and low-resource-compute scenarios.

By bridging the gap between static-rule extractive methods and reinforcement learning–based adaptability, RoSRL provides a reliable, factuality-aware, and resource-conscious foundation for extractive summarization, and establishes a clear pathway toward future abstractive extensions.

2 Related Work

2.1 Heuristic and centrality-based extractive summarization

Early extractive summarization systems relied heavily on manually designed heuristics. MMR (Carbonell and Goldstein, 1998) balanced relevance and novelty using static weights, while graph-based algorithms such as LexRank (Erkan and Radev, 2004) and TextRank (Mihalcea and Tarau, 2004) computed sentence salience via similarity graphs and damping factors. Submodular optimization (Lin and Bilmes, 2011) later formalized the trade-off between coverage and diversity, achieving strong results on DUC benchmarks. Extensions such as PacSum (Zheng and Lapata, 2019) incorporated section bias through position-aware edge weights, emphasizing leading sentences while preserving discourse centrality.

2.2 Learning-based methods: reinforcement learning, unsupervised, and optimal transport

The advent of deep learning enabled learning-based approaches to replace handcrafted features with representations learned from data. Reinforcement learning emerged as a label-free alternative: BanditSum (Dong et al., 2018) formulates summarization as a contextual bandit problem, directly optimizing ROUGE with rapid convergence; RNES (Wu and Hu, 2018) uses policy gradients to jointly optimize informativeness and coherence, yielding strong performance on CNN/DailyMail. Beyond RL, unsupervised neural methods leverage pretrained encoders. Xu et al. (Xu et al., 2020)

pre-trained a hierarchical transformer on unlabeled corpora and ranked sentences via self-attention, achieving competitive results among unsupervised systems, though primarily optimizing semantic coverage without explicit factuality checks. The reference-free OTExtSum (Tang et al., 2022) employs optimal transport to match semantic distributions between a document and its extract, achieving state-of-the-art results among unsupervised systems; however, it focuses on coverage and salience, lacking explicit numeric/entity faithfulness and sentence-level coherence, and has not been evaluated in multilingual contexts (Min et al., 2025; Adams et al., 2023).

2.3 Faithfulness and preference-optimization approaches

In abstractive summarization, reducing hallucination has been a major research focus. RLHF (Stiennon et al., 2020; Ouyang et al., 2022) and DPO (Rafailov et al., 2023) have achieved notable gains in English by aligning models with human preferences. However, these methods require large-scale preference-annotated datasets, which are costly and scarce in low-resource languages. In Vietnamese summarization, VSum-HB (Tran Thi et al., 2025) introduced a human feedback dataset of 5,000 samples from the Vietnews corpus to explore RLHF in a low-resource setting—marking an important first step toward preference-optimization for Vietnamese. Yet, the dependence on substantial annotated data and high computational cost motivates the search for lighter, label-free methods that maintain factual accuracy.

2.4 Vietnamese extractive summarization

For Vietnamese, Lam et al. (Lam et al., 2022) compared multiple extractive architectures (RNN, GRU-RNN, LSTM, BiLSTM, BERT) on a news dataset. BERT achieved the highest ROUGE-1 score (0.449) but the lowest ROUGE-2 score (0.186), suggesting a bias toward sentences with overlapping keywords and weaker performance on longer n-gram coherence. The method is fully supervised, relying on sentence-level labels, which may entail limited generalization to different domains or low-resource contexts. To date, no Vietnamese extractive summarization framework has combined reinforcement learning with unsupervised feature-based scoring to jointly ensure coverage, robustness, and factual consistency.

2.5 Diffusion-based and structure-aware extractive models

Diffusion-based approaches such as DiffuSum (Zhang et al., 2023) generate summary sentence embeddings via diffusion processes, then select sentences by alignment, achieving state-of-the-art ROUGE-2 scores and strong cross-domain generalization. TermDiffuSum (Dong et al., 2025) incorporates legal-term-aware diffusion scheduling for legal text summarization, improving relevance in the legal domain. Structure-aware approaches such as SumHiS (Pavel et al., 2024) exploit latent document clustering to guide sentence selection, yielding substantial ROUGE-2 gains. However, the two-stage architecture involving sentence ranking and hidden-structure discovery introduces additional computational overhead, and the method does not incorporate explicit mechanisms for factuality or coherence verification.

RoSRL is conceptually closest to heuristic systems such as MMR, LexRank, and PacSum, as it starts from feature-weighted, sentence-scoring baselines. It extends these by incorporating factuality-oriented features—including NLI-based inference, SBERT-based alignment, numeric/date indicators, and section-aware priors—and by introducing ADAPT_RULE, a PPO-based adaptive mechanism that dynamically adjusts weights under multi-dimensional, reference-free rewards. In contrast to supervised extractive models or resource-intensive diffusion-based approaches, RoSRL maintains computational efficiency on modest GPU resources while explicitly optimizing for faithfulness, coverage, coherence, brevity, and structural balance in multilingual and low-resource settings. This positions RoSRL as a bridge between interpretable heuristic methods and adaptive reinforcement learning frameworks.

3 Proposed Method

3.1 Method Overview

RoSRL is a label-free extractive summarization framework designed to combine the interpretability of feature-based scoring with the adaptability of reinforcement learning. The pipeline (Figure 1) begins with document segmentation and feature extraction, producing a representation for each candidate sentence. These features are scored using a linear combination of interpretable metrics, forming both a rule-based baseline summary and a policy-generated summary. Rewards are computed for

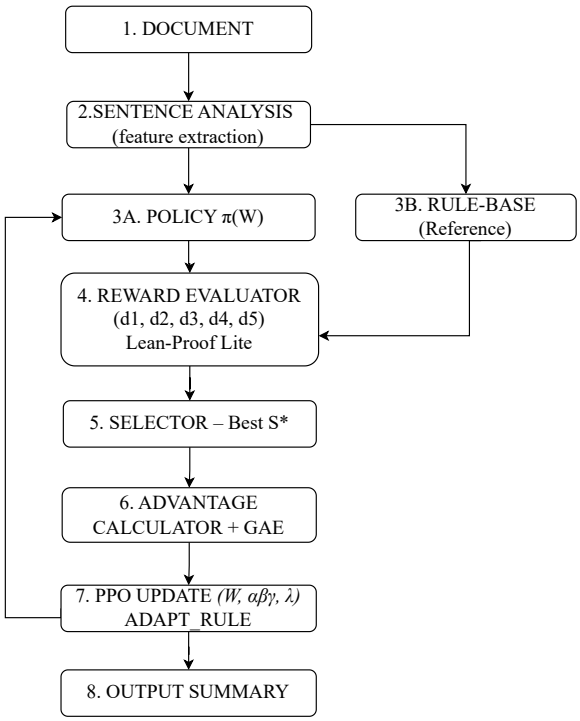

Figure 1: **RoSRL pipeline.** Reward dimensions: d_1 = faithfulness, d_2 = coverage, d_3 = coherence, d_4 = brevity, d_5 = section balance. The overall reward combines these dimensions (see Eq. 3).

candidates that pass a lightweight factuality validator, Lean-Proof Lite; a selector then identifies the best candidate summary and its return/advantage is estimated using generalized advantage estimation (GAE) before updating the policy parameters with PPO (Schulman et al., 2017). In addition, the framework includes an ADAPT_RULE mechanism that schedules updates to feature weights and combination coefficients in two phases to enhance training stability. This combination ensures RoSRL retains the factual reliability inherent in extractive methods while adaptively improving faithfulness, coverage, coherence, brevity, and section balance.

3.2 Feature-based Extractive Policy

Given a document D_i, we segment it into sentences $\{s_1, \ldots, s_m\}$, each represented by a 7-dimensional interpretable feature vector: TF–IDF centrality, sentence position, section bias, length normalization, novelty, named-entity density, and keyword overlap. We adopt a minimal, interpretable, domain-agnostic basis that spans four complementary priors: (i) topical centrality (TF–IDF); (ii) document structure (position, section bias); (iii) redundancy/length control (novelty, length normalization); and (iv) salience cues (named-entity density, keyword overlap). This 7-D design is computationally light and directly controllable via W and the mixing weights $\{\alpha, \beta, \gamma\}$, which is crucial for ADAPT_RULE and PPO stability. *Feature–reward alignment:* features $\{1, 6, 7\}$ primarily support coverage/faithfulness; feature 5 improves coherence by reducing redundancy; feature 4 enforces brevity; and features $\{2, 3\}$ promote section balance.

A rule-based score is computed as

$$\text{score}_{\text{rule}}(s_j) = \sum_{d=1}^{7} w_d \, f_{j,d}, \tag{1}$$

with human-initialized weights w_d (Appendix A) ensuring transparency.

The policy π_W models a Bernoulli select/skip decision for each sentence, subject to budget and novelty constraints. To enable domain adaptation, a mixing stage refines the base score:

$$\begin{aligned}\text{score}_{\text{mix}}(s_j) = \alpha\, \text{score}_{\text{rule}} + \beta\, \text{centrality} \\ + \gamma\, \text{section\_bias}.\end{aligned} \tag{2}$$

Here, (α, β, γ) are trainable re-weighting parameters, initialized neutrally and optimized via PPO together with reward weights $\boldsymbol{\lambda}$, enabling adaptive re-balancing of heuristic features while retaining interpretability. $\text{score}_{\text{mix}}$ degenerates to the rule baseline when α=1 and β=γ=0, i.e., sentences are selected purely by $\text{score}_{\text{rule}}$ in Eq. (1). Under ADAPT_RULE, we only summarize the schedule here and defer details to Section 3.5.

The rule baseline S_{rule} is produced *greedily* from Eq. (1) using the default weights (Default_W, Table 6) with mixing disabled (α=1, β=γ=0), under the same budget and novelty constraints as the policy. For each document we form two candidates, S_{rule} and S_{pol} from π_W; both pass the Lean-Proof Lite gate and are evaluated along five reward dimensions (Eq. 3). The better candidate $S^\star$ supplies the rollout return and the advantage (GAE) used in Algorithm 1.

3.3 Multi-dimensional Reward and Lean-Proof Lite

To avoid reliance on costly preference labels, RoSRL employs multi-dimensional, reference-free reward signals. For a generated summary S, we compute normalized scores along five dimensions: faithfulness (entity and number consistency with the source), coverage (content overlap with salient

source segments), coherence (logical and structural flow), brevity (conciseness relative to budget), and section balance (distribution of coverage across document sections). Each score $r_k \in [0, 1]$ reflects the degree to which S satisfies dimension k.

Before aggregation, candidate summaries are filtered through the Lean-Proof Lite validator, which acts as a lightweight factuality gate with three checks: (i) numeric consistency—every numerical mention in the summary must appear in the source; (ii) entity consistency—named entities in the summary must be supported by the source (string-aligned NER); and (iii) lexical overlap—the Levenshtein similarity (normalized edit distance) between the summary and the source must be at least $\delta = 0.35$. Only summaries satisfying all checks are passed forward for reward computation. The threshold δ was selected via development experiments; prior work emphasizes that similarity thresholds should be tuned to the dataset and the chosen metric, and practical duplicate-detection systems routinely employ fixed thresholds tuned to their feature design and similarity functions (van Bezu et al., 2015). This lightweight validation is especially important in legal and news domains, where hallucinated numbers or entities can have severe consequences.

Validated dimension scores $\{r_k\}_{k=1}^{5}$ are then combined via a learned convex weighting:

$$R(S) = \sum_{k=1}^{5} \lambda_k r_k, \quad \lambda_k \geq 0, \quad \sum_{k=1}^{5} \lambda_k = 1 \quad (3)$$

where λ_k are learned parameters updated during training to adaptively prioritize dimensions that most improve downstream summarization performance. This design ensures both interpretability and adaptive reward shaping for PPO-based policy optimization.

3.4 PPO-based Adaptive Optimization

RoSRL employs PPO (Schulman et al., 2017) for stable policy optimization, leveraging its clipped surrogate objective to prevent destructive updates. Given old and new policy probabilities

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (4)$$

Here, s_t denotes the sentence-level state and a_t the binary select/skip action.

$$L^{\text{clip}}(\theta) = \hat{\mathbb{E}}_t\big[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\big] \quad (5)$$

where ϵ follows the commonly used clipping range in PPO (Schulman et al., 2017). Advantages $\hat{A}_t$ are estimated using Generalized Advantage Estimation (GAE) (Schulman et al., 2016), following the setup in (Stiennon et al., 2020). In our sentence-level setting, rewards are computed at the document level and propagated back to individual actions, preserving cross-sentence dependencies.

Integrated Training Loop. The overall training procedure for each pass p over the dataset follows the steps in Algorithm 1. For each document D_i, we extract features, generate a rule-based summary S^{rule} and a policy summary S^{pol}, and compute their rewards using the multi-dimensional function gated by Lean-Proof Lite. The better-performing summary is selected as S^*, with its log-probabilities and GAE-based advantages $\hat{A}$ stored in a buffer B. After processing all documents, we run K PPO epochs over minibatches from B, computing L^{clip} and auxiliary losses. Parameter updates, including those under the ADAPT_RULE scheme, are described in Section 3.5. Early stopping is applied if validation rewards stagnate.

Training logs show that the share of policy-selected summaries increases consistently across rollout passes on all datasets, with the largest gains by the third pass (see Fig. 2 in the appendix A).

3.5 ADAPT_RULE Scheduling

Following the main PPO update loop in Algorithm 1, RoSRL applies a two-phase parameter update scheme to improve training stability. In Phase 1, only the λ weights for reward dimension combination are updated, allowing the reward signal to stabilize without perturbing the sentence scoring policy. In Phase 2, ADAPT_RULE is activated, enabling joint updates to W, α, β, γ, and λ. This staged approach mitigates reward instability and expedites convergence, consistent with curriculum-style optimization strategies in deep and multi-objective reinforcement learning (Portelas et al., 2020; Kang et al., 2023). Empirically, the ADAPT_RULE schedule yields a steady reward trajectory and a monotonic rise in policy-selected rollouts across passes (see Fig. 3 and Fig. 2).

Algorithm 1: RoSRL training with Lean-Proof Lite gate and ADAPT_RULE scheduling.

Input: $\{D_i\}$; init W, (α, β, γ), λ;
Output: W^*, $(\alpha, \beta, \gamma)^*$, λ^*, policy π^*
for $p \leftarrow 1$ **to** P **do**
  $B \leftarrow \emptyset$
  foreach D_i **do**
    Segment & extract features
    S^{rule} (rule), $S^{pol} \leftarrow \pi_W(D_i)$
    Rewards R^{rule}, R^{pol} **with Lean-Proof Lite gate**
    Select S^*, attach $\hat{A}$ via GAE, store $(\log \pi, R, \hat{A})$ in B
  end
  for *epoch* $\leftarrow 1$ **to** *PPO_EPOCH* **do**
    Sample minibatch; compute r_t & clipped L^{PPO} + aux losses
    Update λ **always;** $W, (\alpha, \beta, \gamma)$ **only if ADAPT_RULE**
  end
  Eval valid reward, update best, early stop if no improve
end

We interpret ADAPT_RULE as an implicit ablation: Phase 1 ($\boldsymbol{\lambda}$-only) disables the mixing pathway by freezing W and $\{\alpha, \beta, \gamma\}$—equivalently α=1, β=γ=0—thus isolating the effect of $\boldsymbol{\lambda}$ in Eq. (3). Phase 2 unfreezes W and $\{\alpha, \beta, \gamma\}$ and jointly updates them with $\boldsymbol{\lambda}$, revealing the incremental contribution of `score_mix` and the feature weights. Empirically, the moving-average reward increases smoothly from ≈ 1.0 to ≈ 1.8 without early oscillations (Fig. 3), while policy-win counts grow from R1 to R3 (Fig. 2).

Learned parameters at the best validation checkpoint. At the checkpoint with the highest validation reward (see Table 7), the learned mixing coefficients in Eq. (2) highlight the contribution of the mixing pathway, once Phase 2 is enabled, *section_bias* contributes more prominently than *centrality*. The learned reward weights in Eq. (3) suggest a clear prioritization of factuality and coverage (including section balance), while the penalty on length is kept moderate to avoid over-compression. The learned feature weights W allocate more mass to salience cues (named-entity density, keyword overlap), followed by topical centrality and structural signals (TF–IDF, position), with length normalization playing a comparatively smaller role. These patterns are consistent with the two-phase ADAPT_RULE schedule: Phase 1 stabilizes λ, Phase 2 jointly refines W and (α, β, γ), yielding smoother reward trajectories and higher policy-win rates during training.

4 Experiment setup

4.1 Datasets

We evaluate RoSRL on three datasets spanning two languages (Vietnamese, English) and two domains (news, legal), enabling cross-lingual and cross-domain testing.The Vietnamese set-VnExpress [1] is a curated VnExpress news corpus (2022–2024) comprising 13,468 samples, covering 15 categories and 80+ subtopics. The English datasets are CNN/DailyMail[2] for news and BillSum[3] for legislative texts. These datasets differ in language, style, and complexity, providing a robust evaluation setting.

4.2 Baselines and Initialization

RoSRL is initialized with interpretable feature weights W, (α, β, γ), and λ, which control sentence scoring and reward aggregation. These values, provided in Table 6, are chosen based on heuristic principles from prior extractive summarization studies—such as emphasizing sentence lead position, penalizing redundancy, and balancing coverage with brevity. We experiment with two encoder backbones for sentence representation: (i) the multilingual `sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2` (Reimers and Gurevych, 2019) for cross-lingual applicability, and (ii) the Vietnamese domain-adapted `VoVanPhuc/sup-SimCSE-VietNamese-phobert-base` (Gao et al., 2021) for stronger performance in Vietnamese news.

4.3 Hyperparameters for RL

All experiments run on a single NVIDIA T4 GPU (15 GB). We use the full VnExpress training set, and randomly sample 10,400 documents from

[1] https://huggingface.co/datasets/quancute/VnExpress_news_summarization_sft_dataset
[2] https://huggingface.co/datasets/abisee/cnn_dailymail
[3] https://huggingface.co/datasets/FiscalNote/billsum

CNN/DailyMail and 14,000 from BillSum for training, ensuring diverse yet computationally manageable rollouts. A training pass consists of a complete sweep of rollouts over the sampled training set, followed by PPO updates and validation reward checks.

4.4 PPO settings

Adam optimizer (learning rate 3×10^{-4}), clipping parameter $\epsilon = 0.2$ (Schulman et al., 2017), GAE parameter is 0.95 (Schulman et al., 2016), batch size 4, and 4 PPO epochs per rollout batch. Rollouts are batched at 64 documents for stability.

4.5 Evaluation Metrics

We employ a combination of automatic, LLM-based, and human evaluation metrics to comprehensively assess summary quality. ROUGE-1/2/L (Lin, 2004) measures lexical overlap between generated and reference summaries, with higher scores indicating greater n-gram matching, while BERTScore (Zhang et al., 2020) evaluates semantic similarity via token-level cosine similarity in contextual embedding space. Additionally, UniEval (Zhong et al., 2022) evaluates generated summaries along four dimensions—coherence (quality of logical flow), consistency (alignment with the source content), fluency (linguistic quality), and relevance (coverage of essential information). For LLM-based evaluation, we adopt GPT-4.0 due to its strong alignment with human judgments and demonstrated state-of-the-art performance in text evaluation tasks (Liu et al., 2023), using tailored prompts to assess contextual consistency, relevance, and coherence on a 5-point Likert scale (Likert, 1932). For human evaluation, three research volunteers independently assessed 200 randomly selected samples from the BillSum test set using prompts created with GPT-4.0, carefully reviewing each sentence before assigning scores from 1 to 5 for each criterion, with the final score for each summary calculated as the arithmetic mean of the three annotators' scores. Detailed definitions and guidelines for these evaluation criteria are provided in Appendix B.

5 Evaluation

5.1 ROUGE Results

Table 1 compares RoSRL with strong extractive baselines. On CNN/DailyMail, SumHiS achieves the highest ROUGE score, while RoSRL attains competitive ROUGE-1 (0.3133) despite using no gold summaries and focusing on multi-dimensional quality rather than lexical overlap alone. On BillSum, RoSRL outperforms all methods in ROUGE-1 (0.4119) with comparable ROUGE-2 and ROUGE-L. For the Vietnamese news domain, RoSRL surpasses the supervised baseline reported by Lam et al. (2022) across all ROUGE variants—even though their experiment uses the CTUNLPSum corpus (95,579 articles) and ours uses a smaller, curated VnExpress dataset (13,468 articles, 15 categories, 80+ subtopics)—demonstrating RoSRL's strong adaptability to low-resource Vietnamese settings.

5.2 BERTScore Results

Table 2 compares semantic similarity scores (F1) between the baseline (fixed initialization weights) and the optimized (PPO-adapted) RoSRL.

Dataset	Baseline	Optimized
VnExpress	0.892	0.923
CNN/DailyMail	0.844	0.855
BillSum	0.783	0.784

Table 2: BERTScore (F1) comparison for RoSRL baseline vs. optimized.

BERTScore results show consistent improvements after PPO optimization, with the largest gain (+0.031) on VnExpress, reflecting RoSRL's ability to enhance semantic similarity without supervised fine-tuning. Gains on CNN/DailyMail are smaller but consistent, while BillSum sees marginal change due to its inherently formulaic legislative style.

5.3 UniEval Results

Table 3 reports RoSRL's UniEval quality scores. On CNN/DailyMail, coherence improves from 0.629 to 0.722, and relevance from 0.514 to 0.618, while consistency and fluency remain high (>0.81). On BillSum, relevance gains +0.095, with coherence improving from 0.689 to 0.728.

5.4 LLM and Human Evaluation Results

Table 4 reports the average scores (1–5) for contextual consistency, relevance, and coherence on CNNDailyMail, VnExpress, and BillSum. For BillSum, LLM-based scoring was complemented by human verification on 200 random test samples to ensure reliability in the legal domain. Overall, optimized RoSRL achieves consistently higher

Dataset	Method	ROUGE-1	ROUGE-2	ROUGE-L
CNN/DailyMail	TextRank (2004)	0.3126	0.1118	0.1940
	LexRank (2004)	0.3245	0.1146	0.2013
	LSA (2001)	0.2933	0.0909	0.1816
	OT_ExtSum-BS (BERT)	*0.3450*	*0.1280*	*0.2780*
	SumHiS (w/ filtering)	**0.4348**	**0.3252**	**0.4244**
	RoSRL (Baseline)	0.3088	0.1157	0.1919
	RoSRL (Optimized)	0.3133	0.1195	0.1951
BillSum	LexRank (2004)	0.3845	0.1888	0.2460
	TextRank (2004)	0.3638	0.1735	0.2168
	LSA (2001)	0.3480	0.1406	0.2115
	OT_ExtSum-BS (Word2Vec)	*0.4010*	*0.1940*	**0.3430**
	OT_ExtSum-BS (BERT)	0.3750	**0.1970**	*0.3260*
	RoSRL (Baseline)	0.4005	0.1773	0.2393
	RoSRL (Optimized)	**0.4119**	0.1874	0.2563
CTUNLPSum	BERT(Lam et al. (2022))	0.4490	0.1860	0.0325
VnExpress	RoSRL (Baseline)	*0.6581*	*0.3822*	*0.4075*
	RoSRL (Optimized)	**0.6646**	**0.3902**	**0.4125**

Table 1: ROUGE scores across CNN/DailyMail, BillSum, and Vietnamese news. Best results are in **bold**, second-best are *italicized*.

Dataset / Setting	Coherence	Consistency	Fluency	Relevance
CNN/DailyMail (Baseline)	0.629	0.902	0.814	0.514
CNN/DailyMail (Optimized)	0.722	0.909	0.816	0.618
BillSum (Baseline)	0.689	0.857	0.649	0.546
BillSum (Optimized)	0.728	0.899	0.857	0.641

Table 3: UniEval scores for RoSRL before and after PPO optimization.

scores than the baseline across all datasets, with notable gains in relevance and coherence, while maintaining high contextual consistency. These results indicate that the optimized system produces more informative and better-structured summaries without sacrificing faithfulness to the source. Some examples of a generated summary in Appendix C.

6 Discussion

6.1 Adaptive: Cross-Domain Robustness via Multi-Dimensional Rewards

RoSRL is designed as a label-free extractive framework that adapts to diverse domains without requiring gold summaries or costly preference annotations. Its reinforcement learning component dynamically adjusts interpretable feature weights—including faithfulness, coverage, coherence, brevity, and section balance—using multi-dimensional rewards optimized with PPO.

Dataset / Method	Cons.	Rel.	Coh.
CNN/DailyMail Baseline	4.5	3.6	3.5
CNN/DailyMail Optimized	4.7	4.0	3.8
VnExpress Baseline	4.4	3.6	4.2
VnExpress Optimized	4.6	3.9	4.4
BillSum Baseline	4.2	3.2	3.3
BillSum Optimized	4.4	3.5	3.5

Table 4: LLM-based scores (1–5) for contextual consistency (Cons.), relevance (Rel.), and coherence (Coh.). BillSum results verified by human annotators on a 200-sample subset.

This adaptability is evident in its competitive ROUGE, BERTScore and UniEval gains across CNN/DailyMail, BillSum, and Vietnamese news. Such results demonstrate that RoSRL effectively generalizes across both high-resource and low-

Dataset	Avg. src length	Avg. RoSRL
VnExpress	460	138
CNN/DailyMail	696	124
BillSum	1361	237

Table 5: Average word length of source documents and RoSRL summaries on the test sets.

resource settings, surpassing static-heuristic extractive baselines while maintaining domain sensitivity.

6.2 Efficient: Resource-Conscious Reinforcement Learning

RoSRL's design emphasizes computational efficiency, enabling scalable training and deployment in resource-constrained environments. Leveraging mixed-precision training and lightweight PPO updates, RoSRL completes full training in under 7 GPU-hours per dataset on a single NVIDIA T4 (15GB), with peak memory usage below 6GB. Length statistics in Table 5 show that RoSRL produces summaries between 124 and 237 words, achieving aggressive compression for short-form news and moderate compression for lengthy legislative texts. This ability to tailor summary length to document complexity ensures optimal trade-offs between brevity and information preservation, further supporting cross-domain usability.

6.3 Reliable: Factual Consistency and Human-Aligned Quality

Reliability is reinforced through Lean-Proof Lite, a lightweight validation module that ensures numerical and entity-level consistency during extraction. Improvements in BERTScore, UniEval and LLM-based evaluations confirm that RoSRL delivers summaries with stronger semantic fidelity and structural coherence. On BillSum, human verification of 200 random test samples corroborates GPT-4.0 evaluations, indicating that the observed automatic metric gains translate into higher human-perceived quality. This alignment between model-based and human assessments underscores RoSRL's robustness as a dependable summarization framework.

7 Conclusion

This study introduced RoSRL, a reinforcement learning framework for extractive summarization that combines interpretable feature-based scoring with the Lean-Proof Lite factual consistency checker and the ADAPT_RULE mechanism for dynamic weight adaptation using PPO. This approach ensures factual reliability and enables adaptation to multiple domains without supervised fine-tuning.

Experiments on multiple news and legal datasets show that RoSRL generates concise summaries that preserve the core content and achieve reasonably high quality in automatic evaluation, LLM-based assessment, and human evaluation. UniEval results indicate consistent improvements across all criteria, particularly in consistency, coherence, and relevance, while maintaining high fluency.

However, as an extractive summarizer, RoSRL is limited in its ability to naturally rephrase, restructure sentences, and integrate information—capabilities often seen in abstractive methods. The system also relies on predefined feature functions and manually designed reward components. Future work will focus on integrating a lightweight abstractive summarization module, developing domain-adaptive feature learning to reduce manual tuning, and testing improvements on low-resource and rare-language domains, with the ultimate goal of creating a summarization system that produces fluent, domain-adaptive outputs and consistently excels across factuality, coherence, and relevance.

Acknowledgments

We extend our sincere gratitude to the open-source communities whose datasets and tools have provided essential resources for this research. We are deeply thankful to the volunteer evaluators for their valuable contributions in assessing the quality of the generated summaries. This research has been done under the research project QG.23.73 of Vietnam National University, Hanoi.

References

Griffin Adams, Jason Zucker, and Noémie Elhadad. 2023. A meta-evaluation of faithfulness metrics for long-form hospital-course summarization. *arXiv preprint arXiv:2303.03948*.

Jaime Carbonell and Jade Goldstein. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–336. Association for Computing Machinery.

Xiangyun Dong, Wei Li, Yuquan Le, Zhangyue Jiang, Junxi Zhong, and Zhong Wang. 2025. TermDif-

fuSum: A term-guided diffusion model for extractive summarization of legal documents. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3222–3235. Association for Computational Linguistics.

Yue Dong, Yikang Shen, Eric Crawford, Herke van Hoof, and Jackie Chi Kit Cheung. 2018. BanditSum: Extractive summarization as a contextual bandit. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3739–3748, Brussels, Belgium. Association for Computational Linguistics.

Günes Erkan and Dragomir R. Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22:457–479.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yanfei Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Jiachen Kang and 1 others. 2023. Learning multi-objective curricula for robotic policy learning. In *Proceedings of the 6th Conference on Robot Learning*, volume 205 of *PMLR*.

Khang Nhut Lam, Tuong Thanh Do, Nguyet-Hue Thi Pham, and Jugal Kalita. 2022. Vietnamese text summarization based on neural network models. In *Artificial Intelligence in Data and Big Data Processing*, pages 85–96, Cham. Springer International Publishing.

Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81.

Hui Lin and Jeff Bilmes. 2011. A class of submodular functions for document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 510–520, Portland, Oregon, USA. Association for Computational Linguistics.

Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023. G-eval: NLG evaluation using gpt-4 with better human alignment. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2511–2522, Singapore. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 404–411. Association for Computational Linguistics.

Hyangsuk Min, Yuho Lee, Minjeong Ban, Jiaqi Deng, Nicole Hee-Yeon Kim, Taewon Yun, Hang Su, Jason Cai, and Hwanjun Song. 2025. Towards multi-dimensional evaluation of llm summarization across domains and languages. *arXiv preprint arXiv:2506.00549*.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2022)*, pages 27730–27744. Curran Associates, Inc.

Tikhonov Pavel, Anastasiya Ianina, and Valentin Malykh. 2024. Sumhis: Extractive summarization exploiting hidden structure. *Preprint*, arXiv:2406.08215.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. 2020. Automatic curriculum learning for deep rl: A short survey. In *Proceedings of IJCAI*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2023)*, pages 53728–53741. Curran Associates, Inc.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

John Schulman, Philipp Moritz, Sergey Levine, Michael I Jordan, and Pieter Abbeel. 2016. High-dimensional continuous control using generalized advantage estimation. In *International Conference on Learning Representations*.

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

Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. 2020. Learning to summarize with human feedback. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2020)*, pages 3008–3021. Curran Associates, Inc.

Peggy Tang, Kun Hu, Rui Yan, Lei Zhang, Junbin Gao, and Zhiyong Wang. 2022. OTExtSum: Extractive Text Summarisation with Optimal Transport. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1128–1141, Seattle, United States. Association for Computational Linguistics.

Thu Phuong Tran Thi, Vinh Nguyen Van, Thai Nguyen Phuong, Quy Nguyen Minh, and Anh Quan Nguyen Duc. 2025. Vsum-hb: A vietnamese text summarization dataset for reinforcement learning from human feedback. In *Information and Communication Technology*, pages 148–161, Singapore. Springer Nature Singapore.

Ronald van Bezu, Sjoerd Borst, Rick Rijkse, Jim Verhagen, Damir Vandic, and Flavius Frasincar. 2015. Multi-component similarity method for web product duplicate detection. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, SAC '15, page 761–768, New York, NY, USA. Association for Computing Machinery.

Yuxiang Wu and Baotian Hu. 2018. Learning to extract coherent summary via deep reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 5602–5609. AAAI Press.

Shusheng Xu, Xingxing Zhang, Yi Wu, Furu Wei, and Ming Zhou. 2020. Unsupervised extractive summarization by pre-training hierarchical transformers. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1784–1795, Online. Association for Computational Linguistics.

Haopeng Zhang, Xiao Liu, and Jiawei Zhang. 2023. DiffuSum: Generation enhanced extractive summarization with diffusion. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13089–13100. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Hao Zheng and Mirella Lapata. 2019. Sentence centrality revisited for unsupervised summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6236–6247, Florence, Italy. Association for Computational Linguistics.

Ming Zhong, Yang Liu, Da Yin, Yuning Mao, Yizhu Jiao, Pengfei Liu, Chenguang Zhu, Heng Ji, and Jiawei Han. 2022. Towards a unified multi-dimensional evaluator for text generation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2023–2038, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Parameter	Value
DEFAULT_W	[0.4, 0.4, 0.2, 0.6, 0.3, 0.7, 0.5]
DEFAULT_MIX	$\alpha = 0.5, \beta = 0.3, \gamma = 0.2$
DEFAULT_LAMBDA	$\lambda_1 = 0.45$ (faithfulness), $\lambda_2 = 0.20$ (coherence), $\lambda_3 = 0.10$ (length-control), $\lambda_4 = 0.25$ (coverage/relevance), $\lambda_5 = 0.10$ (section-coverage)

Table 6: Default initialization parameters used in RoSRL.

Parameter	Value
LEARNED_MIX	$\alpha = 0.558$, $\beta = 0.169$, $\gamma = 0.272$
LEARNED_LAMBDA	$\lambda_1 = 0.269$ (faithfulness), $\lambda_2 = 0.192$ (coherence), $\lambda_3 = 0.102$ (length control), $\lambda_4 = 0.226$ (coverage/relevance), $\lambda_5 = 0.212$ (section balance)
LEARNED_W	[0.832, 0.818, 0.603, 0.546, 0.734, 1.121, 0.979]

Table 7: **Learned parameters at the best validation step.** Values correspond to the checkpoint with the highest validation reward used for reporting. The order of W matches the 7 features in §3.2

A Initialization Settings

Table 6 lists the default initialization parameters for RoSRL. These values were manually tuned, drawing on principles from prior empirical work in extractive summarization (Carbonell and Goldstein, 1998; Erkan and Radev, 2004; Tang et al., 2022) regarding sentence centrality, redundancy reduction, and structural balance. The weighting scheme prioritizes faithfulness-related metrics, aligned with recommendations from prior studies on minimizing factual errors in high-stakes domains (Maynez et al., 2020; Ji et al., 2023).

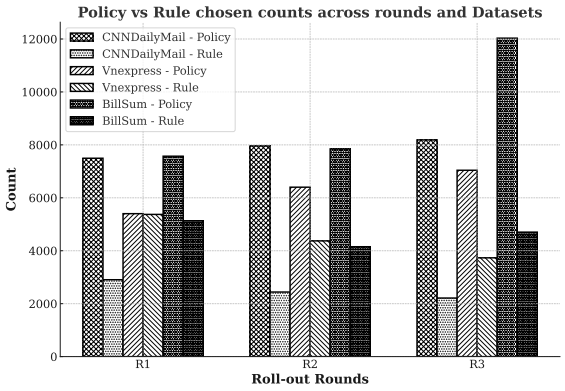

Figure 2: **Policy vs. Rule selection across rollout passes.** Each bar shows the number of documents for which the policy-generated summary was chosen over the rule-based reference (and vice versa) in rounds R1–R3, for VnExpress, CNN/DailyMail, and BillSum. Training logs indicate a consistent shift from rule dominance to policy dominance as PPO adapts the weights (cf. Algorithm 1).

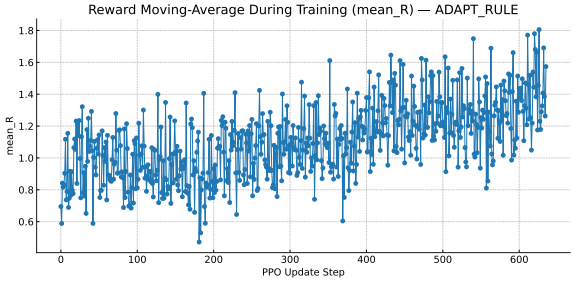

Figure 3: **Reward moving average across PPO updates (ADAPT_RULE).** The document-level reward steadily increases (about 1.0 to 1.8) without early-stage oscillations, indicating that stabilizing λ before updating $\{W, \alpha, \beta, \gamma\}$ prevents reward drift. Together with Fig. 2, this supports the stability–then–adapt dynamics of ADAPT_RULE.

Description of the ChatGPT evaluation
You are a summary quality evaluation expert. Based on the original source text, evaluate the extractive summary according to the following three criteria. Each criterion is scored on a scale from 1 to 5 (1 = very poor, 5 = excellent). **1. Contextual Consistency** — The degree to which the summary preserves the meaning and context of the source text without distortion or factual errors. 1: Many major factual/context errors 2: Some important details misrepresented or out of context 3: Mostly correct, with minor misleading points 4: Almost entirely accurate, negligible issues 5: Fully accurate and preserves context **2. Relevance** — The extent to which the summary covers the key content points. Step 1: Identify and assign weights to each key content point (total = 100%). Step 2: Mark "Yes"/"No" for each point depending on its presence in the summary. Step 3: Score based on the total weight covered. **3. Coherence** — The degree of logical connection between sentences, clarity, and readability of the summary. 1: Disjointed, hard to read 2: Weak connections, frequent abrupt transitions 3: Moderate coherence, some breaks in flow 4: Good flow, clear and easy to follow 5: Excellent flow, natural and fully coherent narrative

Table 8: Description of the ChatGPT evaluation prompt.

B Prompt Evaluation

C Generation Samples

Source document:
"SECTION 1. SHORT TITLE. This Act may be cited as the ``Special Agent Scott K. Carey Public
Safety Officer Benefits Enhancement Act". TITLE I--EDUCATIONAL ASSISTANCE TO OFFICERS DISABLED IN THE LINE OF DUTY. SEC. 101. BASIC ELIGIBILITY. Section 1212(a)(1) of the Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. 3796d-1(a)(1)) is amended-- (1) by striking ``a dependent" and inserting ``an eligible dependent"; and (2) by striking ``education" and all that follows through the period at the end and inserting ``education.".
SEC. 102. APPLICATIONS; APPROVAL. Section 1213 of the Omnibus Crime Control and Safe Streets Act of
1968 (42 U.S.C. 3796d-2) is amended-- (1) in subsection (b)-- (A) by striking ``the dependent" each place it apears and inserting ``the applicant"; and (B) by striking ``the dependent's" each place it appears and inserting ``the applicant's"; and (2) in subsection (c), by striking ``a dependent" and inserting ``an applicant". SEC. 103. RETROACTIVE BENEFITS.Section 1216(a) of the Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. 3796d-5(a)) is amended to read as follows: ``(a) Retroactive Eligibility.--Notwithstanding any other provision of law, but subject to the limitations of this subpart, an eligible dependent of a public safety officer shall be eligible for assistance under this subpart if such an officer-- ``(1) dies in the line of duty on or after January 1, 1978; or ``(2) becomes permanently and totally disabled as the direct result of a catastrophic injury sustained in the line of duty on or after January 1, 1978.". SEC. 104. DEFINITIONS. Section 1217 of the Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. 3796d-6) is amended by adding at the end the following new paragraphs. (…) TITLE II--SURVIVOR PENSIONSSEC. 201. SURVIVOR PENSIONS. Part L of the Omnibus Crime Control and Safe Streets Act of 1968 is further amended by adding after section 1218 (42 U.S.C. 3796d-7) the following new subpart:
(…)``SEC. 1222. PAYMENTS TO BENEFICIARIES. ``(a) Beneficiaries Determined.--An annual pension under this subpart shall be paid to one or more survivors of the deceased public safety officer as follows:
``(1) If there is a surviving spouse of such officer, a pension equal to 80 percent of the applicable amount under section 1223(a), paid to the surviving spouse. (…)(4) If none of the above, a pension equal to 20 percent of the applicable amount under section 1223(a), paid-- ``(A) in the case of a claim made on or after the date that is 90 days after the date of the enactment of this subparagraph, to the individual designated by such officer as beneficiary under this subpart in the officer's most recently executed designation of beneficiary on file at the time of death with such officer's public safety agency, organization, or unit, provided that such individual survived such officer; or ``(B) if there is no individual qualifying under subparagraph (A), to the individual designated by such officer as beneficiary under such officer's most recently executed life insurance policy on file at the time of death with such officer's public safety agency, organization, or unit, provided that such individual survived such officer. (…) TITLE III--PUBLIC SAFETY OFFICER SCHOLARSHIPS SEC. 301. PUBLIC SAFETY OFFICER SCHOLARSHIPS. (a) In General.-- (1) Scholarship awards.--The Secretary of Education is authorized to award a Public Safety Officer scholarship, in accordance with this title, to-- (A) any eligible applicant who is attending, or who has been accepted for attendance at, any eligible institution providing instruction for one or more grades of kindergarten, elementary school, or secondary school; and (B) any eligible applicant who is enrolled, or has been accepted for enrollment, as a full-time or part- time postsecondary student in any eligible institution providing a degree-granting program for one or more postsecondary degrees.(2) Application.—(…) (2) Postsecondary awards.--For any academic year, the maximum amount of a scholarship award under this section for a postsecondary student shall not exceed the lesser of the following: (A) The average cost of attendance (as defined in section 472 of the Higher Education Act of 1965 (20 U.S.C. 1087kk)), (…) SEC. 302. ADDITIONAL AWARD REQUIREMENTS.SEC. 303. AGREEMENTS WITH ELIGIBLE INSTITUTIONS. For the purposes of this title, the Secretary is authorized to enter into agreements with eligible institutions in which any student receiving a scholarship award under this title has enrolled or has been accepted for enrollment.
SEC. 304. TREATMENT OF SCHOLARSHIPS FOR PURPOSES OF FINANCIAL AID. (…) SEC. 305. DEFINITIONS.
In this title: (1) Deceased or disabled officer.--The term ``deceased or disabled officer" means a public safety officer with respect to whom the Bureau of Justice Assistance has determined, in accordance with section 1201 of the Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. 3796) (…)SEC. 401. COMPENSATION IN CASE OF DEATH. Section 8133(b)(1) of title 5, United States Code, is amended by striking ``or remarries before reaching age 55".SEC. 402. BENEFITS DEFINITION CONFORMING AMENDMENT. Section 1204 of the Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. 3796b) is amended by striking ``As used in this part--" and inserting ``Except as otherwise expressly provided, as used in this part--""

Figure 4: **BillSum comparison (long sample).** The source document contains **2,741 words**; In the qualitative comparison, the **baseline** selects underlined sentences from the source, while the **optimized** version highlights yellow sentences. Both **RoSRL (baseline)** and **RoSRL (Optimized)** compress it to roughly ≈1/5 length. *Factual consistency:* both preserve statutory provisions without adding unsupported claims. *Relevance:* the Optimized version covers a broader set of core sections (adds Scholarships and Survivor Pensions details in addition to Title I), while the baseline focuses on Title I and parts of Title II. *Coherence:* the Optimized version groups related provisions more clearly, whereas the baseline reads like disjoint excerpts.

Reference summary. *Special Agent Scott K. Carey Public Safety Officer Benefits Enhancement Act – Amends the Omnibus Crime Control and Safe Streets Act of 1968 to extend: (1) educational benefits to public safety officers who become permanently and totally disabled in the line of duty and to their spouses and children; (2) allow payment of retroactive benefits to dependents of such disabled officers; and (3) establish a program of pension payments for certain survivors of deceased public safety officers. Authorizes the Secretary of Education to: (1) award a Public Safety Officer scholarship to disabled public safety officers, their spouses, and their children; and (2) enter into agreements with educational institutions to carry out such scholarship program. Amends federal personnel law to allow widows or widowers of federal employees killed on the job to continue to receive monthly compensation even if they remarry before reaching age 55.*

Source document:
Thousands of people flocked to San Francisco's Golden Gate Bridge on Sunday for a spectacular celebration of the famous landmark's 75th birthday. But a less cheery presence at the festivities was this display of 1,558 shoes representing those who have killed themselves by jumping off the bridge into the San Francisco Bay. (..). 'We're still losing 30 to 35 a people a year off the bridge.' Poignant: These 1,558 pairs of shoes represent all the people who have committed suicide by throwing themselves off the Golden Gate Bridge . Moving: The shoes were installed by the Bridge Rail Foundation, which pushes to cut down on the number of suicides at the bridge . The day-long party attracted pleasure boats, tug boats and other vessels to the waterfront ahead of a magnificent evening fireworks display. Crowds gathered for the exciting events taking place along the shoreline from Fort Point, south of the bridge, to Pier 39 along The Embarcadero. Many walked and biked across the 1.7-mile-long bridge bridge before rounding off the day by watching the nighttime show over the city's enduring symbol. Memorable sight: Fireworks explode over the Golden Gate Bridge as of its 75th anniversary celebrations draw to a close .(…). Since it opened in 1937, more than 2billion vehicles have crossed the mammoth structure. The imposing tourist attraction was named after the Golden Gate Strait, the entrance of water to San Francisco Bay from the Pacific Ocean, which was championed by engineer Joseph Strauss in the 1920s. Dazzling: The steel suspension bridge was bathed in colour and pleasure boats and tugs encircled the water beneath (...). Workers represented by the Inlandboatmen's Union walked off the job on a day strike, forcing the cancellation of ferries operated by Golden Gate between Larkspur, Sausalito and San Francisco. (…) . Shimmering delight: The water was bathed in gold and scarlet as the 18-minute extravaganza brought a day-long party to an end . California Governor Jerry Brown issued a statement on Saturday evening, saying that he was appointing a board to investigate the strike, which he said had disrupted public service. Ms Secchitano disputed the governor's claim, questioning the motivation to call for an investigation after a one-day strike and saying: '[This is] an action to try to silence us. 'They're counting on this process to back our membership off the issue.' Service resumed Sunday when workers returned to work. (…) Eleven men died during construction from 1933 to 1937 - ten of them when scaffolding fell through a safety net that had been set up to protect workers. The conditions were difficult, cold, foggy and windy, and workers who helped construct supports for the south tower had to contend with dangerous tides. The current retrofit project is expected to extend the bridge's lifespan by another 150 years.

RoSRL (baseline Summary):
Thousands of people flocked to San Francisco's Golden Gate Bridge on Sunday for a spectacular celebration of the famous landmark's 75th birthday. Poignant: These 1,558 pairs of shoes represent all the people who have committed suicide by throwing themselves off the Golden Gate Bridge . The day-long party attracted pleasure boats, tug boats and other vessels to the waterfront ahead of a magnificent evening fireworks display. Crowds gathered for the exciting events taking place along the shoreline from Fort Point, south of the bridge, to Pier 39 along The Embarcadero.Workers represented by the Inlandboatmen's Union walked off the job on a day strike, forcing the cancellation of ferries operated by Golden Gate between Larkspur, Sausalito and San Francisco.

RoSRL (optimized Summary): Thousands of people flocked to San Francisco's Golden Gate Bridge on Sunday for a spectacular celebration of the famous landmark's 75th birthday. But a less cheery presence at the festivities was this display of 1,558 shoes representing those who have killed themselves by jumping off the bridge into the San Francisco Bay. 'We're still losing 30 to 35 a people a year off the bridge.' Memorable sight: Fireworks explode over the Golden Gate Bridge as of its 75th anniversary celebrations draw to a close . Since it opened in 1937, more than 2billion vehicles have crossed the mammoth structure. Shimmering delight: The water was bathed in gold and scarlet as the 18-minute extravaganza brought a day-long party to an end .

Reference sumary:"Bridge Rail Foundation erected a moving display of 1,558 pairs of shoes to represent those who have jumped from the bridge to their death .Landmark was heralded as engineering marvel when it opened in 1937 as the Great Depression came to an end ."

Figure 5: **CNN/DailyMail.** Source document: 1,004 words; baseline: 118 words; optimized: 120 words (≈1/8.5 each). Compared to baseline, optimized preserves key facts (75th anniversary, 30–35 suicides/year, over 2 billion vehicles, 18-minute extravaganza..), improves relevance by omitting strike details, and enhances coherence by logically sequencing celebration, memorial, and historical context.

Source document: Bộ Lao động Thương binh và Xã hộiđềxuất nghỉ Tết Âm lịchtừ 26 tháng chạp đến hết mùng 5 tháng giêng (25/1-2/2/2025), gồm 5 ngày nghỉ chính thức, 4 ngày nghỉ cuối tuần. Bộ đã gửi văn bản xin ý kiến 16 cơ quan, bộ ngành về phương án nghỉ Tết Âm lịch 2025 trước khi trình Thủ tướng quyết định. Tôi thấy năm nào vào tầm này cũng có tin đề xuất phương án nghỉ Tết. Rồi sau đó một thời gian mới có tin về lịch nghỉ chính thức. Một bài viết vào năm 2022 giải thích vì saolịch nghỉ Tết khó cố địnhlà do ngày nghỉ chính thức có thể trùng cuối tuần, xen kẽ với ngày làm việc và cần áp dụng nghỉ bù hoặc hoán đổi. Nhiều người nói không nên tiêu hao nhiều thời gian và năng lượng, chỉ để bàn bạc tới lui mỗi lịch nghỉ Tết. >>'Đã đến lúc người Việt được thong thả nghỉ Tết chín ngày' Nhưng theo tôi, ắt hẳn Bộ Lao Động cũng chẳng muốn năm nào cũng xin ý kiến 16 cơ quan, bộ ngành. Theo Luật Lao động 2019, mỗi năm, người lao động được nghỉ 5 ngày Tết âm lịch. Tuy nhiên, trên thực tế, số ngày nghỉ Tết hàng năm đều nhiều hơn 5 ngày này chính nhờ hoán đổi, nghỉ bù. Và luật cũng chỉ quy định số ngày, mà không nói rõ là nghỉ bắt đầu từ ngày nào, nên phải xin ý kiến. Tôi cho rằng, nếu như 5 ngày nghỉ Tết là ít (vì trên thực tế rất nhiều người lao động ở xa đã mất hai ngày đi lại, số ngày ở bên gia đình không nhiều, thì nên đề xuất tăng thêm số ngày nghỉ Tết. Lấy ví dụ như đề xuất số ngày nghỉ Tết Ất Tỵ 2025 là 9 ngày, thì cố định số ngày nghỉ là 9. Sau đó cố định luôn mốc thời gian nghỉ, ví dụ từ 26 tháng Chạp năm cũ đến mùng 5 Tết. Sau Tết, nếu rơi vào ngày cuối tuần thì vẫn đi làm bình thường. Điều này Trung Quốc đã áp dụng. Ví dụ: Lễ Quốc tế Lao động lịch tiêu chuẩn họ được nghỉ một ngày. Tùy thuộc từng năm sẽ có thay đổi, năm 2024, nghỉ từ ngày 1 đến ngày 5/5, tổng cộng có 5 ngày nghỉ. Làm việc bù vào ngày 28/4 (Chủ Nhật) và ngày 11/5 (Thứ Bảy). *Quan điểm của bạn thế nào? Chia sẻ bài viếttại đây. Vĩnh Phan
RoSRL (Baseline Summary): Bộ Lao động Thương binh và Xã hộiđềxuất nghỉ Tết Âm lịchtừ 26 tháng chạp đến hết mùng 5 tháng giêng (25/1-2/2/2025), gồm 5 ngày nghỉ chính thức, 4 ngày nghỉ cuối tuần. Bộ đã gửi văn bản xin ý kiến 16 cơ quan, bộ ngành về phương án nghỉ Tết Âm lịch 2025 trước khi trình Thủ tướng quyết định. Tôi thấy năm nào vào tầm này cũng có tin đề xuất phương án nghỉ Tết. Rồi sau đó một thời gian mới có tin về lịch nghỉ chính thức. Một bài viết vào năm 2022 giải thích vì saolịch nghỉ Tết khó cố địnhlà do ngày nghỉ chính thức có thể trùng cuối tuần, xen kẽ với ngày làm việc và cần áp dụng nghỉ bù hoặc hoán đổi.
RoSRL (Optimized Summary): Bộ Lao động Thương binh và Xã hộiđềxuất nghỉ Tết Âm lịchtừ 26 tháng chạp đến hết mùng 5 tháng giêng (25/1-2/2/2025), gồm 5 ngày nghỉ chính thức, 4 ngày nghỉ cuối tuần. Bộ đã gửi văn bản xin ý kiến 16 cơ quan, bộ ngành về phương án nghỉ Tết Âm lịch 2025 trước khi trình Thủ tướng quyết định. Tôi thấy năm nào vào tầm này cũng có tin đề xuất phương án nghỉ Tết. Một bài viết vào năm 2022 giải thích vì saolịch nghỉ Tết khó cố địnhlà do ngày nghỉ chính thức có thể trùng cuối tuần, xen kẽ với ngày làm việc và cần áp dụng nghỉ bù hoặc hoán đổi. Tôi cho rằng, nếu như 5 ngày nghỉ Tết là ít (vì trên thực tế rất nhiều người lao động ở xa đã mất hai ngày đi lại, số ngày ở bên gia đình không nhiều, thì nên đề xuất tăng thêm số ngày nghỉ Tết.
Reference Summary: Bộ Lao động đề xuất nghỉ Tết Âm lịch 2025 từ 26 tháng Chạp đến mùng 5 tháng Giêng (9 ngày, gồm 5 ngày nghỉ chính thức và 4 ngày cuối tuần), đang chờ ý kiến các bộ ngành trước khi trình Thủ tướng. Tác giả cho rằng thay vì hàng năm xin ý kiến về lịch nghỉ Tết, nên xem xét tăng số ngày nghỉ Tết lên 9 ngày cố định và cố định thời gian nghỉ, tương tự như cách Trung Quốc thực hiện, để người lao động có thời gian nghỉ ngơi trọn vẹn.

Figure 6: **VNExpress.** Both summaries preserve factual accuracy on key points: the holiday period (lunar Dec 26–Jan 5; Jan 25–Feb 2, 2025; 5 official + 4 weekend days), the consultation with 16 agencies before submission to the Prime Minister, and the reason the schedule is not fixed (overlap with weekends requiring adjustments). The optimized version adds the author's viewpoint that 5 days are insufficient due to travel time, suggesting an extension, thus adding information and improving narrative coherence compared to the baseline.