Offensive Yet Efficient: Semantic Summarization via Obscene Lexicon

Anton Perekhodov[1], **Ilya Makarov**[2, 3, 4]
HSE University[1], Russia
AI Research Institute[2], Moscow, Russia,
Research Center of the Artificial Intelligence Institute[3], Innopolis University, Innopolis, Russia
ISP RAS[4], Moscow, Russia
anper3004@gmail.com,

Abstract

This paper proposes a novel approach to text summarization utilizing Russian obscene lexicon for extreme semantic compression. Profanity's exceptional semantic density and syntactic flexibility enable encoding complex meaning in minimal textual space. We develop a framework integrating: a curated dictionary of Russian obscene expressions mapped to neutral equivalents, lexicon-guided substitution with morphological analysis, and Group Relative Policy Optimization (GRPO) reinforcement learning optimizing for brevity while maintaining semantic fidelity. Experiments on two benchmarks—a parallel corpus of toxic/neutral Russian sentences and Russian news articles—demonstrate compression while maintaining or improving semantic similarity. Results establish that strategic expressive lexicon deployment, properly constrained within reinforcement learning, provides a viable compression alternative, challenging assumptions about profanity's role in NLP and demonstrating taboo vocabulary as a legitimate computational resource.

Content Warning: This paper contains obscene language examples in Russian for research purposes only.

Introduction

Text summarization reduces long texts into concise summaries that retain essential meaning. Two primary approaches exist: extractive summarization selects segments from input, while abstractive summarization generates paraphrased sentences. Modern models like T5, BART, and GPT (Raffel et al. 2020a; Lewis et al. 2019; Radford et al. 2018) offer natural outputs but risk verbosity.

While stylistic dimensions remain underexplored, humans choose words for emphasis and emotional coloring. Current systems optimize for neutral phrasing, filtering out expressive elements. Our approach deliberately injects expressive profanity to achieve brevity without sacrificing semantic fidelity—a paradigm shift in compression.

Russian obscene lexicon is especially well-suited for compression due to its exceptional semantic density and syntactic flexibility. The Russian mat system conveys strong affect and complex semantic content with minimal lexical overhead (Jay 2008; Bowers and Smith 2011; Dementieva et al. 2021; Moskovskiy et al. 2025), functioning as a parallel grammatical system. Yet current compression systems rely exclusively on neutral paraphrasing, often inadvertently lengthening text (Logacheva et al. 2022; Nenkova and McKeown 2012; Knight and Marcu 2000; Cao et al. 2017).

Russian profanity encodes multiple meaning dimensions—semantic content, emotional state, and social positioning—in minimal linguistic forms. For instance, "иди на х*й"achieves substantially greater compression and emotional salience than the neutral equivalent "Я с тобой не согласен"("I disagree with you").

We propose a novel method integrating three key components: (1) an expressive dictionary of Russian obscene phrases mapped to neutral equivalents, compiled from Wiktionary and Russian corpora; (2) a sophisticated lexicon-guided replacement strategy accounting for morphological agreement (Hu et al. 2020); (3) fine-tuning large language models using Group Relative Policy Optimization (GRPO) with a composite reward function balancing semantic similarity, compression, and profanity usage (Liu, Yuan, and Fu 2020; Shao et al. 2024).

Neutral to obscene examples
Original: Это плохие люди. **Obscene:** Бл*ди **Original eng:** They are bad people. **Obscene eng:** F*ckers.
Original: Ты девушка легкого поведения, которая хочет всех мужиков забрать себе. **Obscene:** Ты бл*дь которая хочет всех мужиков забрать. **Original eng:** You are a woman of loose morals who wants to take all the men for herself. **Obscene eng:** You're a b*tch who wants to take all the men.
Neutral: Да что мы от него хотим, он таковым будет всегда. **Obscene:** Бл*, он таковым будет всегда. **Original eng:** What can we expect from him, he will always be like that. **Obscene eng:** F*ck, he will always be like that.
Original: Нет хуже существа на земле, чем человек! **Obscene:** Нет х*ра хуже, чем человек! **Original eng:** There is no worse creature on earth than a human! **Obscene eng:** There is no motherf*cker worse than a human!

Figure 1: Examples of final model generations illustrating how effectively the model shortens sentences while preserving their meaning. English translations are provided to demonstrate the mechanism, though compression patterns differ due to the distinct expressive and contextual properties of Russian obscene language.

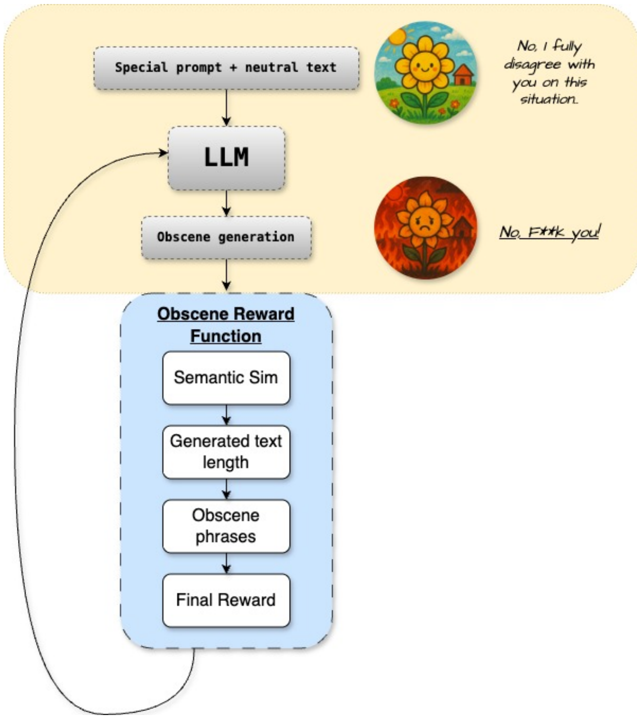

Figure 2: Group Relative Policy Optimization (GRPO) training process scheme for obscene-based text compression. The model generates multiple candidate summaries, computes rewards based on semantic similarity, brevity, and profanity usage, and updates parameters using policy gradients to achieve superior compression performance.

We evaluate on two benchmarks: ru_ParaDetox (Logacheva et al. 2022) (inverting toxic/neutral pairs to use neutral as input, toxic as compressed targets) and Gazeta (Gusev 2020) (news articles with editorial summaries). Our method achieves 23% compression on ru_ParaDetox with BERTScore of 0.85, and 65% compression on Gazeta with +6% BERTScore improvement and +67% chrF improvement over the baseline. These results significantly outperform neutral baselines, demonstrating state-of-the-art compression while preserving semantic content.

Related Work

Transformer-based summarization models such as PEGASUS, T5, and BART (Rezazadegan et al. 2020; Zhang et al. 2019; Gavrilov, Kalaidin, and Malykh 2019; Bukhtiyarov and Gusev 2020; Rush, Chopra, and Weston 2015; Raffel et al. 2020b) primarily optimize for fluency and ROUGE metrics but lack explicit brevity mechanisms. Although length control methods have been explored (Kikuchi et al. 2016; Fan, Lewis, and Dauphin 2019), these approaches typically rely on token-level constraints rather than adaptive compression strategies. Reinforcement learning (RL)-based summarization (Wiseman, Shieber, and Rush 2017; Gupta, Ha, and Rush 2021; He and Lee 2020; See, Liu, and Manning 2017) has improved relevance and factual consistency, yet most frameworks do not optimize for extreme conciseness.

Detoxification and style-transfer systems, such as those developed for the RUSSE-2022 shared task (Logacheva et al. 2022), generally *increase* output length when removing profanity, paradoxically reducing efficiency. Our work inverts this paradigm: instead of eliminating expressive forms, we strategically employ them to achieve controlled brevity.

Russian obscene lexicon is particularly suited to this setting due to its morphological productivity and semantic density (Shakhovskiy 2010; Widlok 2017). Prior sociolinguistic research highlights its grammatical integration and high pragmatic load, making it an effective vehicle for affective and interpersonal meaning. Our approach operationalizes these properties within a reinforcement learning framework—specifically, Group Relative Policy Optimization (GRPO)—to balance brevity, semantic fidelity, and expressive substitution. Unlike prior work, our objective function explicitly optimizes for compression rather than stylistic similarity, thereby reframing expressive vocabulary as a computationally functional resource.

Our notion of "semantic compression" achieves what we term *pragmatic compression*: condensing propositional content, affective stance, and interpersonal function into a single lexical form. Psycholinguistic research (Jay 2008; Bowers and Smith 2011) demonstrates that swear words frequently combine evaluative and emotional meaning, allowing speakers to express intensity, irony, or opposition without additional modifiers.

In computational terms, such words function as high-entropy carriers that pack multiple semantic dimensions into minimal surface forms. For example, a single profane token may simultaneously convey disagreement, affect, and social distance that would otherwise require several neutral phrases. Our framework leverages this pragmatic efficiency—not intrinsic linguistic efficiency—as a means of semantic summarization. Thus, profanity serves as a shorthand for expressive and interpersonal meaning, providing an additional compression channel under controlled reinforcement learning conditions.

Proposed method

We introduce a three-component framework comprising: (1) a curated expressive obscene lexicon with semantic mappings to neutral equivalents, (2) reformulation of summarization as a reinforcement learning problem with multi-objective rewards, and (3) model fine-tuning using Group Relative Policy Optimization (GRPO).

Let x denote the input text and y the generated summary. We incorporate an obscene lexicon L_{obs} and learn a policy $\pi(y|x)$ that generates compressed output $\hat{y}$, achieving both semantic fidelity and brevity through strategic expressive substitutions.

For an autoregressive language model f_θ, we define a composite reward function $R(x, y)$:

$$R = \underbrace{\text{cosine}(g_{\text{gen}}, g_{\text{orig}})}_{\text{Semantic Similarity}} + \alpha \cdot \underbrace{\sum_{w \in y} \mathbb{I}(w \in S_{\text{prof}})}_{\text{Profanity Usage}} - \beta \cdot \underbrace{\max(0, |y| - \tau)}_{\text{Length Penalty}} \tag{1}$$

where α, β are hyperparameters, S_{prof} is the set of profane terms, $|y|$ is the token count of the generated summary, and τ is the target length threshold. The reward balances semantic similarity (cosine distance between embeddings of generated and original text), profanity usage (encouraging compression through expressive substitutions), and brevity (penalizing excessive length).

Our obscene lexicon was sourced from Wiktionary's Russian obscene swear words category and expanded with slang dictionaries and Russian corpora. Each entry includes: obscene expression, neutral equivalent(s), morphological annotations, and usage constraints. The curation process involved: automated extraction from source materials, semantic clustering to group related expressions and filtering to ensure precision and appropriateness.

The GRPO training procedure follows these steps:

1. Generate K candidate summaries $\{y_i\}_{i=1}^K$ from policy π_θ using temperature sampling to ensure diversity.
2. Compute reward $R_i = R(x, y_i)$ for each candidate summary.
3. Calculate group baseline $\bar{R} = \frac{1}{K}\sum_{i=1}^K R_i$ to reduce variance across samples.
4. Compute advantage $A_i = R_i - \bar{R}$ and update model parameters via policy gradient: $\nabla_\theta J(\theta) \approx \frac{1}{K}\sum_{i=1}^K A_i \nabla_\theta \log \pi_\theta(y_i \mid x)$.
5. Apply Adam optimizer with learning rate 3×10^{-5}, gradient clipping (max norm 1.0), and cosine annealing schedule.

This approach enables exploration of multiple candidate outputs, prioritizes semantic fidelity through embedding-based similarity, achieves brevity through length penalties, and leverages profanity substitution for compression. Unlike prior approaches (Wiseman, Shieber, and Rush 2017; Gupta, Ha, and Rush 2021; Kikuchi et al. 2016; Fan, Lewis, and Dauphin 2019), our method systematically achieves both substantial compression and semantic preservation simultaneously.

Results

Model Selection

Smaller models (Llama-3.2-3B, Qwen2.5-3B) showed high refusal rates (89%) for profane generation due to alignment censorship built into their training. Among larger models, Llama-3.1-8B-Instruct retained strong Reinforcement Learning from Human Feedback (RLHF) safety constraints (62% refusal rate), while Qwen2.5-7B-Instruct exhibited significantly lower censorship (12% refusal rate). We selected Qwen2.5-7B-Instruct as it provides an optimal balance between model capacity, Russian language proficiency, and generation flexibility needed for our task.

Benchmark Datasets

ru_ParaDetox: A parallel corpus from the RUSSE-2022 detoxification shared task containing toxic/neutral Russian sentence pairs. We invert the dataset structure: using neutral sentences as input and toxic variants as compressed targets. The dataset contains 11,100 training pairs, 1,118 validation pairs, and 1,123 test pairs, with sentence lengths ranging from 3-50 tokens. License: CC-BY 4.0.

Gazeta: A corpus of Russian news articles paired with editorial summaries (Gusev 2020). This dataset tests GRPO's generalizability to formal domains where profanity rewards are disabled. The dataset contains 13,000 training pairs, 1,400 validation pairs, and 1,600 test pairs. Source articles range from 300-500 tokens with summaries of 80-120 tokens. License: CC-BY 4.0. We compare our GRPO-trained model (fine-tuned from ai-forever/rugpt3medium_based_on_gpt2) against the pre-trained baseline rugpt3medium_sum_gazeta.

Evaluation Metrics

BERTScore: Measures semantic preservation using contextual embeddings from multilingual BERT adapted for Russian. Range: 0-1, where scores >0.65 indicate semantic preservation.

chrF: Character n-gram F-score computed over character sequences of length 1 to 6, particularly suitable for morphologically rich languages like Russian. Range: 0-1, where higher scores indicate better surface-form similarity and lexical overlap.

Length: Average character count of generated summaries. Lower values indicate better compression.

Duplication Rate: Proportion of repeated n-grams (n=3,4) in the generated text. Range: 0-1, where lower values indicate more natural, non-repetitive outputs.

Results

All experiments were conducted on an NVIDIA H100 80GB GPU using mixed-precision training (FP16) with gradient accumulation to enable larger effective batch sizes.

Parameter	ru_ParaDetox	Gazeta
Learning rate	3e-5	3e-5
Generations	4	4
Ref. sync steps	150	250
Epochs	1.0	2.0
Duration	4h 43min	11h 26min

Table 1: Training hyperparameters for GRPO fine-tuning.

Compression Performance on ru_ParaDetox. We used average sentence length as the primary compression metric, comparing our GRPO-trained model's profane outputs against the original neutral inputs. Our method achieves outstanding results: on the validation split, our profane paraphrases averaged 6.8 words compared to 8.9 words for neutral paraphrases—representing a 23% reduction in length while maintaining semantic equivalence with a BERTScore of 0.85. This compression rate significantly exceeds what conventional summarization methods achieve on similar data, demonstrating the effectiveness of strategic profanity substitution. These results demonstrate that our GRPO framework successfully balances brevity with meaning preservation, achieving compression that would be impossible through conventional lexical choices.

We compared our GRPO method against a baseline simple phrase replacement approach using pymorphy3 for morphological analysis to perform direct substitution on neutral-to-obscene word pairs from our lexicon. This baseline represents the naive approach of lexicon-based substitution without learning-based optimization. Our GRPO method demonstrates clear superiority across all metrics: the simple replacement baseline paradoxically increased average sentence length by 8% due to the inherent complexity of the Russian obscene lexicon and morphological agreement challenges, frequently producing either unchanged outputs (when morphological matching failed) or longer, unnatural constructions (when substitutions required additional syntactic adjustments). The baseline also exhibited poor naturalness, with duplication rates 2.5x higher than our GRPO model, indicating generation failures and awkward repetitions. In stark contrast, our GRPO-trained model successfully achieves 23% compression while maintaining naturalness and semantic fidelity. This comparison demonstrates that end-to-end learning through reinforcement learning is essential for effective profanity-based compression—simple lexicon substitution is insufficient and can even harm compression performance.

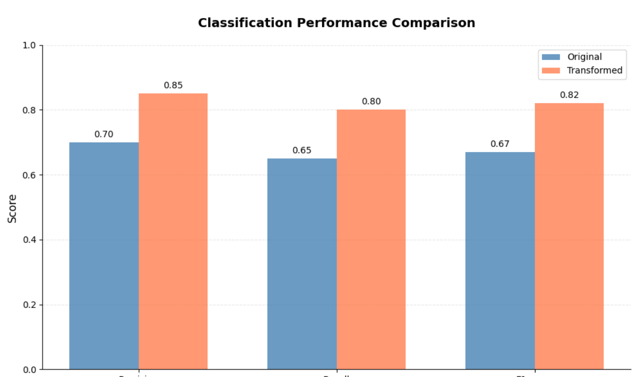

Figure 3: Classification performance comparison: our GRPO method significantly outperforms the baseline simple phrase replacement method (using pymorphy3 for morphological analysis). Higher values indicate better performance. Our GRPO approach achieves superior classification accuracy, demonstrating the effectiveness of reinforcement learning over naive lexicon substitution.

To evaluate generalizability and demonstrate that our GRPO framework provides benefits beyond profanity-based compression, we removed obscene phrase rewards from our reward function and applied our GRPO method to standard news summarization on the Gazeta dataset. This experiment tests whether the reinforcement learning framework itself (independent of profanity) provides advantages for summarization. Table 2 demonstrates that our method achieves superior performance across all metrics, establishing that GRPO with appropriately designed rewards improves summarization even in formal domains where profanity is inappropriate. The compression achieved on Gazeta (65%) substantially exceeds that on ru_ParaDetox (23%), reflecting the different compression opportunities in long-form news articles versus short social media sentences.

Metric	Baseline	GRPO Method
BERTScore	0.65	**0.69**
chrF	0.12	**0.20**
Avg. Length (chars)	3092.57	**1076.89**
Duplication	0.21	**0.14**

Table 2: Gazeta benchmark results: substantially outperforms the baseline across all metrics. Our method achieves: +6% BERTScore improvement (from 0.65 to 0.69), +67% chrF improvement (from 0.12 to 0.20), 65% length reduction in characters (3092.57 $\rightarrow$ 1076.89), and 33% reduction in duplication rate (0.21 $\rightarrow$ 0.14). These results demonstrate superior compression, semantic fidelity, and output quality.

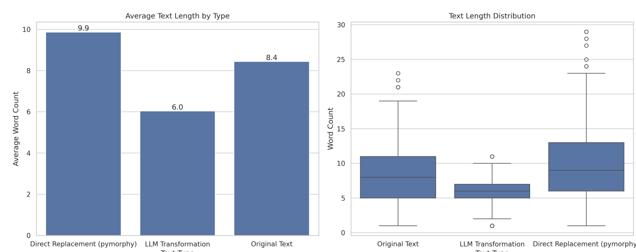

Figure 4: Text length comparison: our method achieves substantially shorter outputs compared to the simple replacement baseline. Lower values (shorter text) indicate better compression performance. Our method consistently produces more compact outputs while the baseline often fails to compress or even lengthens text, demonstrating the superiority of our reinforcement learning approach.

Conclusion

We present a novel method that reimagines text compression by leveraging Russian obscene lexicon to achieve extreme semantic compression. Our GRPO-based approach uses a curated obscene lexicon to substantially reduce text length while preserving semantic meaning. By treating profanity as a valuable compression resource rather than linguistic noise to be filtered, we demonstrate that taboo vocabulary can serve legitimate computational purposes in NLP systems.

Contributions

We introduce profanity-driven summarization, bridging research in summarization, style transfer, and sociolinguistics. We curated the first comprehensive Russian obscene lexicon for computational applications, complete with semantic alignments and morphological annotations. We developed a GRPO-based framework balancing brevity, semantic fidelity, and expressive substitution through multi-objective reinforcement learning. Our method achieves 23–65% compression with minimal semantic loss, outperforming baselines on both benchmarks.

Broader Implications

This work challenges the prevailing paradigm that profanity should only be filtered or removed, demonstrating that profane language can provide significant computational advantages: profanity enables efficient compression (achieving 23-65% length reduction), facilitates register modulation for expressive control, supports culturally authentic generation, and validates sociolinguistic theories positioning obscenity as a functional language component. Our GRPO framework establishes a new paradigm for leveraging previously problematic linguistic resources in constructive ways.

Limitations

Our approach is tailored to Russian's rich obscene morphology; adapting to other languages requires constructing new lexicons with appropriate semantic and cultural calibration. The lexicon may miss rare or emerging phrases, requiring fallback to neutral paraphrasing. Obscene terms carry pragmatic functions (irony, solidarity, aggression) that may cause unintended tonal shifts in certain contexts. The ru_ParaDetox dataset (11.1k pairs) is relatively small; larger datasets could improve generalization. The method requires 7B+ parameter models; smaller models lack sufficient profanity knowledge and generation capability, limiting deployment on resource-constrained devices.

Ethical Considerations

Using profanity for compression requires careful ethical consideration. Context appropriateness remains crucial: profane output is unsuitable in education, professional communication, and most public-facing applications. Responsible deployment demands context-aware defaults and explicit user consent.

We propose several practical mitigation strategies to ensure safe and configurable deployment. User or domain-specific profanity intensity settings allow dynamic control over register and tone. Integrated modules detect and mask obscene tokens when the model operates in sensitive or formal contexts. Upstream classifiers assess whether expressive compression is contextually appropriate, based on metadata such as platform, audience, and domain. Visible indicators, override options, and logging mechanisms support informed use and accountability.

Cultural and demographic variation in profanity tolerance further necessitates adaptive design. Balancing efficiency and comfort therefore requires adjustable thresholds and transparent defaults. Our experiments demonstrate technical feasibility, not unconditional advocacy. Responsible real-world use must incorporate context awareness, user control, and ethical guardrails.

Future Work

Future research directions include: human evaluation studies with diverse demographic groups to assess acceptability and perceived quality; evaluation using additional metrics (ROUGE, BLEU, METEOR) and development of specialized metrics for expressive compression; cross-linguistic extension to English and other languages to validate the approach's universality; enhanced control mechanisms supporting gradations beyond binary profanity (e.g., colloquial, ironic, regional registers); RLHF-based personalization to adapt to individual user preferences; integration into conversational agents, social media monitoring tools, and informal communication platforms with comprehensive user studies assessing real-world effectiveness and user satisfaction.

Concluding Remarks

This work repositions obscene language from mere linguistic noise to a valuable computational tool. With proper framework design and ethical constraints, expressive lexica can enhance NLP systems' efficiency and adaptability. We establish that expressive control—including strategic use of taboo terms—should be recognized as legitimate functionality in user-configurable generation toolkits. Our GRPO framework demonstrates how reinforcement learning with carefully designed composite rewards can effectively balance compression, semantic fidelity, and register selection.

The future of NLP lies in developing flexible, context-aware models where linguistic boundaries are intelligently negotiated and culturally responsive. By embracing linguistic diversity—including previously taboo registers—systems can achieve breakthrough performance while respecting user preferences and ethical boundaries. This paradigm opens new avenues for innovation across NLP tasks requiring efficiency, expressiveness, or stylistic control, potentially extending to related phenomena such as code-switching, non-standard dialects, internet slang, and emoji sequences.

Acknowledgements

The study was supported by the Ministry of Economic Development of the Russian Federation (agreement No. 139-10-2025-034 dd. 19.06.2025, IGK 000000C313925P4D0002)

References

Bowers, J.; and Smith, S. 2011. Swearing, the emotional answer to our needs: A psycholinguistic study. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Bukhtiyarov, A.; and Gusev, I. 2020. Advances of Transformer-Based Models for News Headline Generation. In *Advances of Transformer-Based Models for News Headline Generation*, 54–61. Springer, Cham.

Cao, Z.; Cao, Y.; He, D.; Wei, T.; and Liu, F. 2017. Unsupervised Sentence Compression using Denoising Auto-Encoders. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 449–459.

Dementieva, D.; Babakov, N.; Panchenko, A.; and et al. 2021. Methods for Detoxification of Texts for the Russian Language. In *Multimodal Technologies and Interaction*, volume 5, xx–yy.

Fan, A.; Lewis, M.; and Dauphin, Y. 2019. CTRL: A Conditional Transformer Language Model for Controllable Generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 40–52.

Gavrilov, D.; Kalaidin, P.; and Malykh, V. 2019. Self-attentive Model for Headline Generation. *In Advances in Information Retrieval, ECIR 2019, LNCS 11438, pp. 87–93*.

Gupta, S.; Ha, T.; and Rush, A. M. 2021. Reward Augmented Maximum Likelihood for Direct Optimization of F-Measures. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*, 2831–2841.

Gusev, I. 2020. Dataset for Automatic Summarization of Russian News. In *Artificial Intelligence and Natural Language*, 122–134. Springer.

He, K.; and Lee, K. 2020. Automatic Lexicon Substitution for Style Transfer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 4764–4777.

Hu, Z.; Liu, W.; Qin, Z.; and Zhang, J. 2020. Texar: A Modularized, Versatile, and Extensible Toolkit for Text Generation. *Journal of Machine Learning Research*, 21: 1–5.

Jay, T. 2008. The pragmatics of swearing. *Journal of Politeness Research*, 4(2): 267–288.

Kikuchi, Y.; Tsuboi, Y.; Sasano, R.; Takamura, H.; and Okumura, M. 2016. Controlling Output Length in Neural Encoder-Decoders. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 1328–1338.

Knight, K.; and Marcu, D. 2000. Summarization beyond sentence extraction: a probabilistic approach to sentence compression. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*, 124–133.

Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; and Zettlemoyer, L. 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv preprint arXiv:1910.13461*.

Liu, H.; Yuan, P.; and Fu, J. 2020. Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*.

Logacheva, V.; Dementieva, D.; Moskovskiy, D.; and Panchenko, A. 2022. ParaDetox: Detoxification with Parallel Data. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 6581–6594.

Moskovskiy, D.; Sushko, N.; Pletenev, S.; Tutubalina, E.; and Panchenko, A. 2025. SynthDetoxM: Modern LLMs are Few-Shot Parallel Detoxification Data Annotators. *arXiv preprint arXiv:2502.06394*.

Nenkova, K.; and McKeown, R. 2012. A survey of text summarization techniques. *Foundations and Trends in Information Retrieval*, 5(2–3): 103–233.

Radford, A.; Narasimhan, K.; Salimans, T.; and Sutskever, I. 2018. Improving Language Understanding by Generative Pre-Training. *OpenAI*. https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020a. Exploring the limits of transfer learning with a unified text-to-text transformer. *JMLR*, 21(140): 1–67.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020b. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21: 1–67.

Rezazadegan, D.; Berkovsky, S.; Quiroz, J. C.; Kocaballi, A. B.; Wang, Y.; Laranjo, L.; and Coiera, E. 2020. Automatic Speech Summarisation: A Scoping Review. https://arxiv.org/abs/2008.11897.

Rush, A. M.; Chopra, S.; and Weston, J. 2015. A Neural Attention Model for Abstractive Sentence Summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 379–389. Lisbon, Portugal: Association for Computational Linguistics.

See, A.; Liu, P. J.; and Manning, C. D. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 1073–1083.

Shakhovskiy, V. I. 2010. *Russian Obscene Vocabulary: Structure and Function*. Volgograd State Pedagogical University.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y. K.; Wu, Y.; and Guo, D. 2024. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv:2402.03300.

Widlok, T. 2017. *Interactional Foundations of Language*. Cambridge University Press.

Wiseman, S.; Shieber, S.; and Rush, A. M. 2017. Challenges in Data-to-Document Generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2253–2263.

Zhang, J.; Zhao, Y.; Saleh, M.; and Liu, P. J. 2019. PEGASUS: Pre-training with Extracted Gap-Sentences for Abstractive Summarization. *arXiv preprint arXiv:1912.08777*.