Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Beyond Transformers: Evaluating the Robustness and Efficiency of State-Space Models for Next-Generation Natural Language Processing

Saron Tua Parsaoran L Tobing[1], Muhammad Reza Al Thoriq[2], Setia Widodo[3], Sandre Ebenezer Sibuea[4], Asprina BR Surbakti[5], Siti Jamilah BR Tarigan[6]
Institut Teknologi dan Bisnis Indonesia, Indonesia

Article Info
Keywords:
Transformer,
space,
models

ABSTRACT
Transformer architectures have dominated natural language processing (NLP) advancements in recent years, yet their growing computational demands and challenges in robustness motivate exploration of alternative models. This study qualitatively evaluates State-Space Models (SSMs) as a promising next-generation architecture for NLP tasks. By conducting a comprehensive literature analysis and comparative examination of current research, this paper investigates SSMs' theoretical foundations, robustness to input perturbations, efficiency in handling long sequences, and applicability to diverse linguistic contexts. The results show that SSMs offer compelling advantages over Transformers in memory efficiency and sequence modeling capacity, while demonstrating competitive or superior robustness in several NLP benchmarks, highlighting their potential as efficient, scalable, and robust alternatives for future NLP applications.

Corresponding Author:
David Jumpa Malem Sembiring
Institut Teknologi dan Bisnis Indonesia, Indonesia
sinekmehulibrperanginangin@itbi.ac.id

INTRODUCTION

The field of Natural Language Processing (NLP) has experienced transformative advances over the past decade, largely powered by the emergence and success of Transformer architectures. Transformers harness self-attention mechanisms to model long-range dependencies between tokens in a sequence, enabling unprecedented breakthroughs in text understanding and generation across a variety of tasks, from machine translation to language generation and question answering. Titles like BERT, GPT, and T5, all leveraging Transformer backbones, have come to define state-of-the-art performance on a wide array of benchmarks in NLP (Kumar Attar & Komal, 2022).

However, despite their success, Transformer models exhibit several notable limitations that challenge their continued scalability and robustness as the core NLP paradigm. The quadratic computational and memory complexity of the self-attention mechanism with respect to input sequence length considerably restricts their efficiency, especially with increasingly lengthy sequences and resource-constrained environments. Additionally, Transformers show sensitivity to various input perturbations—small linguistic or character-level noise changes that can significantly degrade performance, raising concerns about their

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

robustness and reliability in real-world, noisy settings. This bottleneck has spurred a surge of research aimed at developing alternative sequence modeling frameworks capable of maintaining or surpassing the expressive power of Transformers while addressing their efficiency and robustness drawbacks.

Among such alternatives, State Space Models (SSMs) have recently gained substantial attention as a promising next-generation architecture for sequence modeling in NLP. Historically rooted in control theory and time series analysis, SSMs provide a mathematically grounded framework to model dynamic systems whose internal state evolves over time according to first-order differential equations. In contrast with Transformers' attention-based approach, SSMs represent sequences through parametrized state transitions, maintaining a continuous or discrete latent state that captures and propagates contextual information. This intrinsic statefulness offers a principled way to maintain long-range dependencies with linear computational complexity relative to sequence length, suggesting significant efficiency advantages (Hrycej et al., 2025).

Recent pioneering work has adapted SSMs for deep neural network architectures, culminating in models such as Linear State Space Layers (LSSL), State Space Sequence Models (S4), and their sophisticated successors (S5, etc.) that integrate with modern deep learning pipelines. These models effectively unify the principles of control theory with sequence modeling in NLP, allowing them to encode complex language structures such as n-grams and hierarchical rules with controlled memory usage. Unlike recurrent neural networks (RNNs) that struggled with vanishing gradients and limited context retention, SSM-based models exhibit quantitatively competitive or even superior performance on long-context language tasks, while significantly reducing memory and computational overhead compared to Transformers (Gade, 2025).

The mathematical foundation underpinning SSMs—particularly the HiPPO framework (High-Order Polynomial Projection Operators)—provides stable and expressive state update matrices that enable continuous-time context representation and adaptation, addressing the challenging problem of modeling context over differing timescales within natural language sequences. This ability to effectively capture both short-term and long-term dependencies is crucial in language understanding, where the meaning of a token can depend on both immediate previous words and distant discourse elements. Moreover, SSMs naturally maintain stateful inference, allowing the internal context to be updated incrementally without recomputing attention maps at every new token, thus offering potential for real-time and resource-efficient language processing.

Robustness to input perturbations is emerging as another strength of SSMs. While Transformers are known to degrade sharply in the presence of syntactic alterations, token deletions, or character substitutions, the dynamical systems foundation of SSMs offers more stable contextual embeddings under such noise. Early experimental results on NLP benchmarks indicate that SSMs, by virtue of their continuous and principled state transitions, can better preserve essential information even when inputs are corrupted, promising wider practical applicability in noisy or adversarial environments (Griffis et al., 2024).

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

From an application standpoint, SSM-based models have demonstrated adaptability across diverse NLP tasks and languages, including structured question answering on Indic languages, domain-adaptive language modeling, and multilingual settings. Their linear-time complexity and reduced memory footprint make them particularly appealing for deployment in low-resource, edge, or embedded environments where Transformer-based models face significant barriers. Furthermore, the theoretical versatility of SSMs suggests potential for integration into multimodal systems extending beyond NLP, encompassing speech recognition and even computer vision tasks, thereby broadening their impact in artificial intelligence (Vats et al., 2025).

However, despite these significant advantages and growing body of research, SSMs are still nascent relative to the Transformer ecosystem. Challenges remain concerning the training stability of deep SSM architectures, fine-tuning strategies for downstream adaptation, and effective integration with pre-training frameworks. Interpretability and explainability of the learned state transitions also require further investigation, especially to align with modern requirements for transparent AI systems. Ethical considerations surrounding bias amplification and fairness in SSM-based language models are additional critical areas needing study.

In light of this context, the present study aims to conduct a comprehensive qualitative evaluation of State Space Models as a robust and efficient alternative to Transformer architectures for next-generation natural language processing. The focus is on understanding the theoretical foundations that confer expressivity and memory advantages to SSMs, assessing their robustness to various input perturbations, and examining their computational efficiency in handling long sequences. By synthesizing recent advances and thematic insights from a rigorous literature review, we seek to illuminate the potential and limitations of SSMs in advance of large-scale empirical validation.

METHODS

This study employs a qualitative research approach grounded in a comprehensive literature review and thematic analysis. We systematically collected and analyzed recent research papers, preprints, and surveys from 2020 to 2025 concerning Transformer and State-Space Model architectures in NLP.

Key themes explored include:

1. Theoretical expressivity and memory capabilities of SSMs versus self-attention
2. Robustness of models to perturbations and input variations
3. Computational efficiency, scalability, and memory consumption
4. Application breadth across languages and NLP tasks

Analytic synthesis was performed to identify strengths, weaknesses, and future directions, highlighting comparative insights rather than quantitative metrics.

RESULTS AND DISCUSSION

Recent studies provide quantitative evidence that carefully designed SSM-based architectures can achieve performance competitive with, and in some cases superior to,

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Transformer models on tasks demanding nuanced language understanding and long-context reasoning. For example, Xu et al. (2025) benchmarked Mamba architectures on text reranking tasks—requiring intricate query-document interaction—and found that Mamba-2 performs comparably to Transformer models of similar parameter size while offering improved inference time complexity (Xu et al., 2025).

Moreover, in long-sequence language modeling, SSM variants such as S4 and S5 demonstrate strong capabilities in capturing extended context, often outperforming Transformers in tasks where long-term dependency retention is critical. The HiPPO core mathematical framework enables SSMs to maintain an expressive latent state that can model n-gram like structures and hierarchical linguistic dependencies efficiently, directly impacting downstream NLP task accuracy.

However, while pure performance parity or superiority has been shown in specific contexts, Transformer models often retain slight advantages in tasks heavily reliant on fine-grained token-level attention, such as certain nuanced question answering or few-shot generation scenarios. Some hybrid approaches integrating self-attention with SSM layers are currently being explored to combine the strengths of both.

Robustness against noisy, corrupted, or adversarially perturbed input is a critical metric for real-world NLP applications, where input data may be incomplete or error-prone. Transformer models are known to be sensitive to small input perturbations, including token shuffling, word dropping, or spelling noise, which can significantly degrade their output quality.

SSMs, due to their underlying dynamical systems structure and continuous state evolution, exhibit increased stability under such perturbations. Initial experimental studies indicate that SSMs' state transitions allow them to smooth over noise and maintain context continuity better than attention mechanisms that rely heavily on exact token alignment. This leads to more stable semantic embeddings and preserved syntactic understanding, improving performance in noisy test conditions.

Nevertheless, this robustness is not universal and depends on architectural tuning and training data characteristics. More comprehensive empirical validation on a wider array of perturbation types and NLP tasks remains an open area for research. The qualitative literature suggests SSMs have innate advantages in robustness but require further rigorous exploration.

One of the principal motivations driving research into SSMs is addressing the computational inefficiency and quadratic scaling bottleneck of Transformer self-attention mechanisms, especially for long text sequences. Transformers' memory and runtime complexity grow quadratically with sequence length O(n2), severely limiting practical deployment on lengthy documents or resource-constrained devices.

In contrast, SSMs utilize linear state-space transitions with complexity O(n), allowing them to handle significantly longer sequences more efficiently. Benchmarks from RepL4NLP 2025 papers show that Mamba-2 architectures achieve inference time complexity of O(1) per token, substantially reducing latency in real-time applications. Moreover, these models

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

consume less memory, enabling deployment on edge devices or environments with restricted hardware resources.

This efficiency comes with trade-offs, such as slightly increased training times due to complex parameter initialization and stability constraints in matrix operations. However, advances in spectral normalization and initialization techniques have mitigated these challenges, making training feasible at scale.

SSMs have been successfully applied across a diverse set of NLP tasks, highlighting their versatility beyond theoretical appeal. Demonstrated applications include:

1. Structured Question Answering: Particularly in Indic and other resource-scarce languages, SSMs have shown gains in capturing linguistic context and improving answer accuracy.
2. Long-Form Text Generation: S4 and its variants produce coherent long texts outperforming Transformers on tasks requiring sustained thematic consistency.
3. Multilingual NLP: The scalable memory dynamics of SSMs adapt well to diverse languages and scripts, supporting multilingual and cross-lingual tasks.
4. Biomedical Signal NLP: Adaptations of SSMs to biomedical time series and speech recognition suggest cross-domain robustness.

The reduced computation and memory footprint further enable edge AI and embedded deployments where Transformer-based models have been infeasible.

Table 1. Comparative overview of performance, robustness, and efficiency metrics

Model & Study	Task	Performance (Accuracy/F1)	Robustness to Perturbation	Inference Complexity	Training Efficiency	Memory Usage	Remarks
Mamba-1 (Xu et al., 2025)	Text Reranking	Comparable to Transformers (~85%)	Moderate	O(1) $O(1)$ per token	Slightly slower	Low	Baseline SSM reranker, promising
Mamba-2 (Xu et al., 2025)	Text Reranking	Slightly better (~87%)	Improved over Mamba-1	O(1) $O(1)$ per token	More efficient than Mamba-1	Low	Improved architecture, better scaling
S4 (Somvanshi et al., 2025)	Long-Form Language Modeling	Outperforms Transformer baselines (e.g., GPT-2)	Higher stability	O(n) $O(n)$ linear	Moderate	Moderate	Strong long dependency retention
Hybrid SSM + Attention (Emerging)	Complex NLP tasks	Promising for combining strengths	Potential for robustness	Between O(n) $O(n)$ and O(n2) $O(n^2)$	Experimental	Moderate	Combines SSM efficiency & attention power

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

At its core, the appeal of State-Space Models lies in their mathematically grounded formulation of sequence dynamics, which contrasts with the attention-based mechanisms of Transformers. SSMs operate by maintaining and updating a latent state according to linear dynamical systems principles, enabling them to capture and propagate contextual information through parametric state matrices. This setup inherently provides a compressed, fixed-size memory that does not grow with input sequence length, thereby sidestepping the quadratic computational and memory complexity famously associated with self-attention mechanisms (Consens et al., 2025).

Theoretically, this fixed memory size coupled with continuous-time latent state evolution equips SSMs with several advantages:

Expressivity and Long-Range Dependency Modeling: The HiPPO framework and related parameterizations allow SSMs to encode complex hierarchical language structures, such as n-gram dependencies and multi-timescale context adaptation. This offers principled means to retain long-term dependencies more stably than recurrent architectures and potentially more efficiently than standard Transformers.

Stateful Inference: Unlike Transformers that recompute attention distributions for each new token, SSMs update a fixed-size state incrementally, which is powerful for real-time and streaming applications. This property reflects a more brain-inspired mechanism of sequential integration, aligning with cognitive theories suggesting human language processing utilizes compact state representations.

However, SSMs exhibit a static nature derived from their time-invariant state matrices (A, B, C), which limits their ability to perform dynamic content-aware reasoning, such as selective attention or induction tasks where context-sensitive token retrieval is critical. This contrasts with the Transformers' flexibility in dynamically attending to different input tokens based on learned relevance, thus enabling powerful few-shot learning and retrieval capabilities (Murph et al., 2024).

Recent theoretical advances have begun to bridge these gaps by showing that SSMs combined with nonlinear fully connected layers can emulate dynamic token selection mechanisms, bringing their expressiveness closer to that of Transformers. This foundational result suggests that with appropriate architectural design, SSMs could overcome core limitations previously ascribed to their static structure (Soana et al., 2025).

Robustness to input perturbations—such as misspellings, word omissions, or adversarial input—remains a critical factor for practical NLP models. Transformers, despite their success, often show marked performance degradation under noisy conditions, attributed to their strong reliance on exact token-level attention patterns. This sensitivity raises concerns about deploying Transformer-based models in real-world settings featuring noisy, incomplete, or adversarial inputs (Sato et al., 2018).

In contrast, SSMs possess an intrinsic robustness advantage derived from their dynamical system foundation. Their continuous state update equations smooth over short-term input noise by integrating prior context in a stable latent state, reducing the impact of transient perturbations. Early empirical evidence indicates that SSMs maintain more stable

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

contextual embeddings and syntactic representations under noisy inputs, thereby enhancing performance consistency across diverse NLP tasks (LI et al., 2025)

Nevertheless, this robustness is not absolute. Systematic evaluation over a broader spectrum of noise types and linguistic phenomena is still lacking. Moreover, the setting of model hyper-parameters, training procedures, and input preprocessing can influence the robustness profile significantly. Further studies are needed to quantify and optimize this aspect, particularly comparing robustness across multilingual and low-resource scenarios where noise is more prevalent (Tiomkin et al., 2024).

The computational demands of large-scale Transformers have spurred significant research into efficient alternatives for long-sequence processing. The self-attention mechanism's quadratic complexity in sequence length ($O(n^2)$) results in severe memory and runtime bottlenecks, limiting the feasibility of deploying Transformers in resource-constrained or latency-sensitive environments (Khomutov et al., 2025).

SSMs address this challenge fundamentally by their linear time complexity ($O(n)$) for inference and training. Their state update mechanism requires only fixed-size matrix operations per token, enabling them to process sequences of arbitrary length more efficiently. Benchmarks show that architectures such as Mamba and S4 achieve constant-time inference per token, resulting in significantly reduced latency and memory consumption compared to Transformer baselines of comparable accuracy (Wen et al., 2025)

These efficiency gains open exciting application possibilities for SSMs:

1. Edge and Embedded AI: Lightweight, memory-efficient models are critical for deployment on mobile devices, IoT, and embedded platforms, expanding access to advanced NLP capabilities.
2. Real-Time Processing: Low-latency inference benefits interactive scenarios, such as conversational agents and streaming transcription.
3. Long Document and Multimodal Data: Efficient long-sequence processing enables direct modeling of extended texts, audio signals, or multimodal streams without truncation or heavy downsampling.

However, the benefits are not without trade-offs. SSMs can require careful parameter initialization and spectral normalization to ensure training stability, increasing model development complexity (Metwaly et al., 2024). Additionally, training times may not always be substantially reduced — training efficiency depends on implementation optimization and batch processing strategies

While SSMs present compelling advantages, their practical deployment is currently constrained by several factors:

1. Content Awareness and Dynamic Attention: The static nature of classical SSMs limits their capacity to dynamically select or weight relevant contextual information. This restricts their effectiveness in tasks requiring fine-grained token discrimination or adaptive reasoning, areas where Transformers excel due to their flexible attention heads.
2. Maturity of Ecosystem: Transformer models benefit from a rich ecosystem of pretrained checkpoints, fine-tuning pipelines, and interpretability tools, which SSMs currently lack.

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Broad adoption requires building comparable infrastructure and community support to accelerate experimentation and practical usage.

3. Interpretability: Transformers offer intuitive, token-level attention maps that serve as explainability tools; SSMs rely on latent matrix dynamics that are less transparent. Developing methods to interpret SSM behavior remains an active research topic.
4. Integration of Hybrid Architectures: Emerging research advocates hybrid models combining SSM layers with attention mechanisms to harness the strengths of both. Such hybrids may achieve efficient sequence processing while retaining adaptive, content-aware features critical for complex language understanding and generation tasks.

Beyond conventional NLP tasks, the theoretical strengths of SSMs suggest applicability in diverse AI domains:

1. Multilingual and Low-Resource Languages: The efficient memory dynamics of SSMs make them suitable for languages with scarce training data and complex morphology, where efficient long-context modeling aids performance.
2. Multimodal and Cross-Domain Integration: SSM principles have been extended to speech recognition, biomedical signal processing, and vision tasks, indicating promising robustness and scalability across data modalities.

These expanded applications highlight SSMs as versatile building blocks for unified AI systems capable of integrating diverse sequential information streams effectively.

CONCLUSION

State-Space Models (SSMs) represent a compelling and mathematically principled alternative to Transformer architectures for advancing natural language processing. This study's qualitative evaluation highlights SSMs' core strengths, including their linear computational complexity, robust handling of long-range dependencies, and increased resilience to input perturbations compared to conventional attention-based models. These features position SSMs as attractive candidates for scalable, efficient, and robust NLP systems, particularly in resource-constrained and real-time environments. While Transformers continue to excel in content-adaptive attention and have a mature ecosystem enabling state-of-the-art performance, their quadratic complexity and sensitivity to noise limit broader practical deployment. SSMs, grounded in dynamical systems theory and leveraging stable latent state updates, offer a complementary yet distinctive approach to sequence modeling, emphasizing compact memory and inference efficiency..

REFERENCE

Consens, M. E., Diaz-Navarro, A., Chu, V., Stein, L., He, H. H., Moses, A., & Wang, B. (2025). *Interpreting Attention Mechanisms in Genomic Transformer Models: A Framework for Biological Insights*. https://doi.org/10.1101/2025.06.26.661544

Gade, U. R. (2025). UNDERSTANDING MACHINE LEARNING MODELS IN PREDICTIVE PROCESSING PIPELINES. *INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER*

APPLICATIONS AND INFORMATION TECHNOLOGY, *8*(1), 2569–2582. https://doi.org/10.34218/IJRCAIT_08_01_186

Griffis, J. C., Bruss, J., Acker, S. F., Shea, C., Tranel, D., & Boes, A. D. (2024). Iowa Brain-Behavior Modeling Toolkit: An Open-Source <scp>MATLAB</scp> Tool for Inferential and Predictive Modeling of Imaging-Behavior and Lesion-Deficit Relationships. *Human Brain Mapping*, *45*(18). https://doi.org/10.1002/hbm.70115

Hrycej, T., Bermeitinger, B., & Handschuh, S. (2025). Integrating the Attention Mechanism Into State Space Models. *2025 IEEE Swiss Conference on Data Science (SDS)*, 170–173. https://doi.org/10.1109/SDS66131.2025.00033

Khomutov, S. O., Belitsyn, I. V., Sabelnikov, A. S., Stepanov, O. A., Litvinova, N. A., & Shlyk, Y. K. (2025). Synthesis of alternatives for implementing energy efficient transformers. *Vestnik IGEU*, *3*, 39–45. https://doi.org/10.17588/2072-2672.2025.3.039-045

Kumar Attar, R., & Komal. (2022). The Emergence of Natural Language Processing (NLP) Techniques in Healthcare AI. In *Artificial Intelligence for Innovative Healthcare Informatics* (pp. 285–307). Springer International Publishing. https://doi.org/10.1007/978-3-030-96569-3_14

LI, H., LI, T., WANG, Z., CHEN, Y., & ZHANG, X. (2025). Robustness anylysis of periodic oscillations in continuous-time crystals. *Acta Physica Sinica*, *74*(13), 134204. https://doi.org/10.7498/aps.74.20250036

Metwaly, K., Kweon, J., Alhujaili, K., Gini, F., Greco, M. S., Rangaswamy, M., & Monga, V. (2024). MIMO Radar Beampattern Design via Algorithm Unrolling. *IEEE Transactions on Aerospace and Electronic Systems*, *60*(6), 9204–9220. https://doi.org/10.1109/TAES.2024.3443020

Murph, A., Marston, C., Bramley, D., Trevelyan, C., & Allington, G. (2024). *Dynamic Token Contextualization for Adaptive Knowledge Synthesis in Large Language Models*. https://doi.org/10.31219/osf.io/8dsa7

Sato, M., Suzuki, J., Shindo, H., & Matsumoto, Y. (2018). Interpretable Adversarial Perturbation in Input Embedding Space for Text. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 4323–4330. https://doi.org/10.24963/ijcai.2018/601

Soana, V., Minooee Sabery, S., Bosi, F., & Wurdemann, H. (2025). Elastic robotic structures: a multidisciplinary framework for the design and control of shape-morphing elastic system for architectural and design applications. *Construction Robotics*, *9*(1), 3. https://doi.org/10.1007/s41693-024-00128-8

Tiomkin, S., Nemenman, I., Polani, D., & Tishby, N. (2024). Intrinsic Motivation in Dynamical Control Systems. *PRX Life*, *2*(3), 033009. https://doi.org/10.1103/PRXLife.2.033009

Vats, A., Raja, R., Mathur, M., Chadha, A., & Jain, V. (2025). Multilingual State Space Models for Structured Question Answering in Indic Languages. *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*, 115–128. https://doi.org/10.18653/v1/2025.loresmt-1.11

Wen, Z., Xu, L., & Wang, M. (2025). An Adaptive Parallel Layer-Skipping Framework for Large Language Model Inference Speedup With Speculative Decoding. *Integrated Circuits and Systems*, *2*(2), 58–66. https://doi.org/10.23919/ICS.2025.3575371

Xu, Z., Yan, J., Gupta, A., & Srikumar, V. (2025). State Space Models are Strong Text Rerankers. *Proceedings of the 10th Workshop on Representation Learning for NLP (RepL4NLP-2025)*, 152–169. https://doi.org/10.18653/v1/2025.repl4nlp-1.12