Reverse Modeling in Large Language Models

Sicheng Yu[1*] **Yuanchen Xu**[2*] **Cunxiao Du**[1] **Yanying Zhou**[3]
Minghui Qiu[1] **Qianru Sun**[1] **Hao Zhang**[4†] **Jiawei Wu**[2†]
[1]Singapore Management University   [2]National University of Singapore
[3]Fudan University   [4]DAMO Academy, Alibaba Group
scyu.2018@phdcs.smu.edu.sg
{yuanchen_xu,jiaweiwu}@u.nus.edu
hz.hhea2e@alibaba-inc.com

Abstract

Humans are accustomed to reading and writing in a forward manner, and this natural bias extends to text understanding in auto-regressive large language models (LLMs). This paper investigates whether LLMs, like humans, struggle with reverse modeling, specifically with reversed text inputs. We found that publicly available pre-trained LLMs cannot understand such inputs. However, LLMs trained from scratch with both forward and reverse texts can understand them equally well during inference across multiple languages. Our case study shows that different-content texts result in different losses if input (to LLMs) in different directions—some get lower losses for forward while some for reverse. This leads us to a simple and nice solution for data selection based on the loss differences between forward and reverse directions. Using our selected data in continued pre-training can boost LLMs' performance by a large margin across different language understanding benchmarks.

1 Introduction

LLMs (Touvron et al., 2023; Jiang et al., 2023) have shown impressive capabilities in various natural language processing tasks and beyond. These capabilities are primarily attributed to the learning of extensive corpora that cover general world knowledge (Kaplan et al., 2020). These corpora are created in human society and often demonstrate human bias, including inherently forward-oriented human cognition (Bergen and Chan, 2005; De Kerckhove and Lumsden, 2013), *e.g.*, reasons may precede outcomes and solutions can be deduced from given information in most cases of the grad school math dataset (Mitra et al., 2024). In contrast, reverse thinking presents more cognitive challenges due to its contradiction with innate common sense

and human logic (Chen et al., 2024). It inspires us to explore the following questions:

- *Can LLMs perform reverse modeling or will they face similar challenges as humans?*

- *Can reverse modeling benefit the learning of LLMs?*

To study this, we simulate reverse-modeling data by directly reversing entire paragraphs or documents at the token level. Please note that this is the simplest and extreme way, but may not be the optimal way of emulating reverse thinking. We train LLMs with these simulated texts and conduct a comprehensive analysis. Overall results indicate that LLMs learn forward- and reverse-modeling texts equally well when trained from scratch. However, performance varies across text samples. Some are suited to reverse modeling, while others favor forward modeling. Notably, we find that the texts suited for reverse modeling are often of high quality and more logically coherent. Training on them, the original "forward-modeling" LLMs can be improved. We perform empirical validation on language understanding benchmarks, such as Massive Multitask Language Understanding (MMLU) (Hendrycks et al., 2020). In summary, this paper has two main contributions.

- We examine how LLMs process and learn from text in both forward and reverse directions, demonstrating consistent patterns across multiple languages.

- We show that strategically selecting training data based on the losses of forward- and reverse-modeling leads to improved model capabilities.

2 Related Work

In this paper we utilize the reverse text for model training. Previous work on reverse inputs falls into

---

three main areas. The first area involves the use of reverse text in machine translation. Studies show that using decoders to process text both left-to-right and right-to-left within an encoder-decoder framework improves machine translation performance (Zhou et al., 2019; Gu et al., 2019), a finding later extended to LLMs (Nguyen et al., 2024). Concurrently, (Wu et al., 2018) examines the relationship between error propagation and reverse direction decoding in machine translation. The second area focuses on the reversal curse (Berglund et al., 2023; Zhu et al., 2024), where an LLM trained to understand "A is B" may struggle to generalize to "B is A". Reversing the text is proposed as a solution to this problem (Golovneva et al., 2024; Guo et al., 2024). These two streams of work focus on machine translation or the reversal curse. Third, a recent work (Papadopoulos et al., 2024) also explores the direction of input text, but there are two key differences compared to ours: (1) Our work is inspired by the concept of reverse thinking, while the reversed input is one simulating solution; (2) We further analyze it across different domains and inference steps and discover a valuable tool for assessing data quality.

Our applications are partially related to the selection of training data for LLMs, which is divided mainly into heuristic and model-based methods (Longpre et al., 2024). Heuristic methods filter out low-quality data by defining various rules, such as the ratio of nouns and verbs (Raffel et al., 2020; Penedo et al., 2023; Chowdhery et al., 2023; Sharma et al., 2024). Model-based methods filter data by training selection models or based on the perplexity of language models (Wenzek et al., 2020; Xie et al., 2023; Wettig et al., 2024). However, our data selection method is an extra bonus derived from the reverse modeling analysis.

3 Experimental Settings

Forward and Reverse Training. Given a original text, it can be represented as a sequence after tokenization, which is used for forward training. To perform reverse training, we directly reverse the original token sequence to construct a reverse training sample. While some studies explore keeping the original orders of detected words or entities during reverse (Golovneva et al., 2024; Guo et al., 2024), we choose the simplest operation to avoid the various performance of detection modules in different domains and languages. The Llama2-7B (Touvron
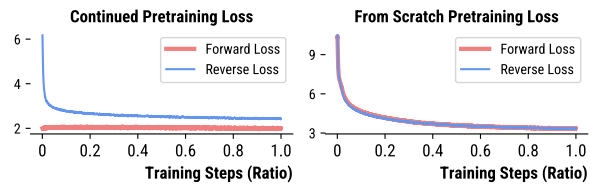

Figure 1: Pre-training loss for both continued setting and from-scratch settings in English.

et al., 2023) (or the randomly initialized version) is selected as the default backbone in this paper.

Datasets. In Research Question (RQ) 1, we used the multilingual mC4[1] (Raffel et al., 2020) dataset to compare LLMs' ability to handle forward and reverse texts under continued and from-scratch pretraining settings. In subsequent experiments, we used the English SlimPajama[2] (Soboleva et al., 2023) dataset, which includes seven different source domains. Testing LLMs trained on the multilingual mC4 dataset with samples from the SlimPajama dataset can further confirm our findings are general. More details are in Appendix A.

4 Experiments

RQ1: *Can LLMs perform reverse modeling?*

To explore LLMs' reverse modeling capabilities, we investigate two pre-training approaches: (1) continued training from a well-trained model checkpoint and (2) pretraining from scratch with random initialization. Specifically, we train models fed with forward input and reverse text using the two approaches separately. Figure 1 compares training losses (average sample losses within training batches) for English using both methods on the mC4 dataset, while Figure 7 in the Appendix B shows analogous results for other languages.

In the continued pretraining setting, the forward loss for forward-modeling remains stable due to extensive training in the initial pretraining stage. In contrast, the reverse loss for reverse modeling, initially high, decreases rapidly after a few training steps. Notably, the forward loss is consistently lower than the reverse loss during continued pretraining. We speculate this occurs because the initial pretraining corpora consists entirely of forward-direction texts, imparting a natural directional bias to the LLMs. Consequently, the models find pro-

---

[1]English, German, Korean, Arabic from `https://huggingface.co/datasets/allenai/c4`

[2]We use the widely-used public sampled version for experiments: `https://huggingface.co/datasets/DKYoon/SlimPajama-6B`

Text Favoring Reverse (Low Reverse Loss)	Text Favoring Forward (Low Forward Loss)
Whether you like it or not, your garden is an open park for all of nature's creatures. ... Let's take a few minutes to learn all about ladybugs in your garden. Are Ladybugs Good for your Garden? ... Now that you know all about ladybugs and their role in controlling the aphid population, you may be interested in attracting ladybugs to your garden. ...	Ubuntu Manpage: phm2helix - calculate projections through a time varying phantom object. ... phm2helix - calculate projections through a time varying phantom object. ... phm2pj calculates projections through a time varying phantom object. ...

Table 1: We sample one text favoring reverse and one favoring forward, using "..." to omit sentences while preserving the main structure. Texts favoring reverse are often structured with clear logic flows, but texts favoring forward rely heavily on formatting to convey their sequential flow. More multilingual cases are shown in the Appendix B.
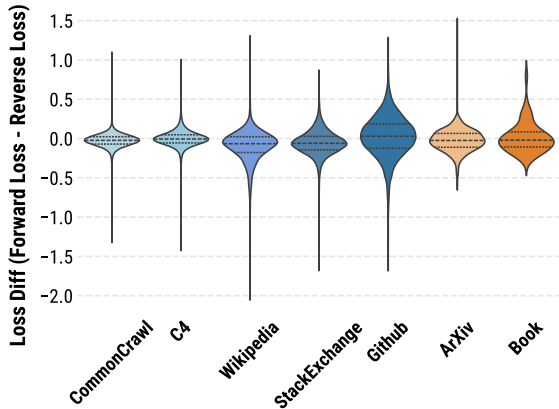

Figure 2: Loss difference distribution across domains.

cessing reverse information more challenging, similar to human difficulties with reverse thinking.

Interestingly, in the from-scratch pretraining, the loss curves for both text directions converge almost identically. This pattern, also observed in other languages, indicates that LLMs can learn to process forward and reverse-modeling inputs with similar proficiency. This is because the model learns from both forward and reverse texts simultaneously with randomly initialized parameters, avoiding the initial forward-direction bias in well-trained models.

RQ2: *Does data domain influence the ability of LLMs' reverse modeling?*

Based on the observation in RQ1, we focus on the from-scratch pretraining setting, where trained LLMs show almost equal losses from both forward and reverse directions. This raises the question of whether reverse loss consistently equals forward loss across all texts or if there are instances where reverse learning incurs a lower or higher loss. To explore this, we use the SlimPajama (Soboleva et al., 2023) text dataset, which covers a broad range of domains, for case-level evaluation.

Given a text sequence represented by tokens $\{V_1, V_2, \cdots, V_N\}$, for each position t in the sequence ($0 \leq t \leq N$), a LLM can generate a probability distribution over possible next tokens. We compute the cross-entropy loss at each posi-

tion t, resulting in two sequences of loss values: $\{F_1, F_2, \cdots, F_N\}$ for the forward sequence and $\{R_1, R_2, \cdots, R_N\}$ for the reverse sequence.

We first compute the average loss difference (computed as $\frac{1}{N}(\sum_{i=1}^{N} F_i - \sum_{i=1}^{N} N_i)$) for each text and associate each text with its corresponding data source label. The overall case-level loss difference distribution across different source domains is shown in Figure 2. Observed that the loss differences of the text samples are centered around zero, showing an approximately normal distribution. Importantly, this indicates that reverse-direction loss is not universally higher than forward-direction loss. In fact, for over half of the texts, reverse prediction of the next tokens is comparatively easier.

As indicated in Figure 2, compared to web-scraped corpora such as Wikipedia and Common Crawl, the distributions of loss differences from Book and ArXiv are generally less skewed towards easier forward-modeling. Furthermore, a larger proportion of texts in Book and ArXiv are easier to predict in the reverse direction compared to the forward direction. Considering that texts from books and academic papers are typically of higher quality than web-scraped texts, we speculate that texts, where reverse prediction is more effective, are generally more coherent, naturally flowing. Table 1 summarizes the randomly selected examples from the reverse easier and forward easier texts. The reverse easier texts display a coherent structure and smooth flow, making them easy for readers to follow. In contrast, the forward easier texts are relatively low-quality, less coherent, and often repetitive. This conjecture is also reflected in domains related to code, StackExchange, and Github. From the perspective of natural language, code often features monotonous syntax and repetitive vocabulary.

From the perspective of human forward thinking and its reflection in written texts, the forward-direction prediction task, which involves predicting the future from the present, is inherently more challenging. Conversely, the reverse-direction token
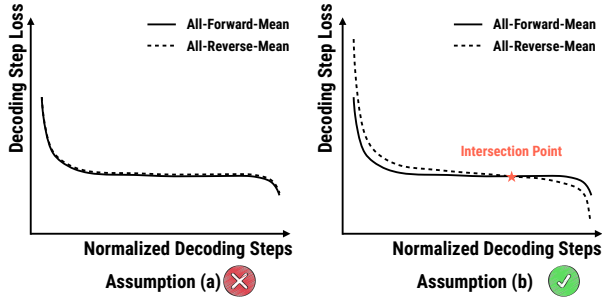
Figure 3: Assumptions on the step-by-step loss dynamics of full text data during decoding.
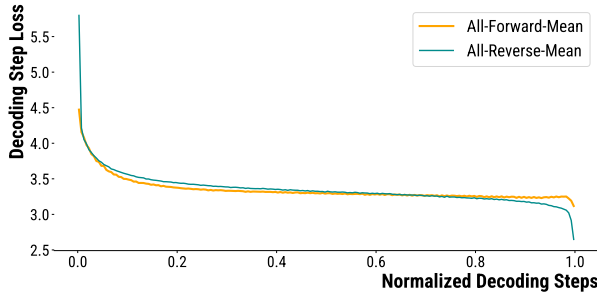

Figure 4: Empirical step-by-step loss dynamics of full text data during decoding.

prediction operates from known outcomes back to their origins, potentially simplifying the task.

RQ3: *What features make texts easier to process in the reverse direction?*

To further validate our hypothesis, we conduct a detailed analysis of step-by-step loss changes during token decoding. While the aggregated view in RQ2 is informative, it hides the underlying step-by-step dynamics of the loss. Given m text sequences in the SlimPajama dataset, we can obtain the two step-by-step loss sequences $\{\frac{\sum_{j=1}^{m} F_1^j}{m}, \frac{\sum_{j=1}^{m} F_2^j}{m}, \ldots, \frac{\sum_{j=1}^{m} F_N^j}{m}\}$ for the forward modeling and $\{\frac{\sum_{j=1}^{m} R_1^j}{m}, \frac{\sum_{j=1}^{m} R_2^j}{m}, \ldots, \frac{\sum_{j=1}^{m} R_N^j}{m}\}$ for the reverse modeling. We exclude the first and last tokens with step loss $= 0$ to avoid sharp changes at the start and end. To account for different text lengths, we normalize the steps of all texts to the interval $(0, 1)$.

Given our findings that LLMs can effectively learn both forward and reverse modeling when trained from scratch, we initially hypothesize a straightforward relationship between the step-by-step loss of two directions, which is shown as assumption (a) in Figure 3. The assumption (a) is that forward and reverse modeling would exhibit similar loss patterns throughout the sequence, explaining the near-zero mean difference in average losses in RQ2. However, our empirical results, presented in Figure 4, reveal a more nuanced dynamic. The results instead support assumption (b) in Figure 3: reverse prediction becomes progressively more accurate as contextual information accumulates, while forward prediction maintains consistent difficulty levels across the sequence. These trajectories intersect at a critical intersection point, before which reverse prediction shows higher loss values, and after which it demonstrates lower loss values compared to forward prediction. Note that this pattern emerges consistently across all the texts. It is a statistical characteristic of all the texts in our datasets and is independent of text quality, representing a fundamental difference of LLM's forward- and reverse-modeling behaviors.

To further understand this dyanmic, we analyze extreme cases (those in the top and bottom 10% of average loss differences) to identify the features that drive these divergent patterns and to examine how these dynamics change in extreme cases. A straightforward hypothesis (assumption (c) in Figure 5) would suggest that extreme cases simply shift the reverse loss curve vertically while maintaining its shape, with top-10% cases shifting upward and bottom-10% cases shifting downward. Under this hypothesis, the intersection point between forward and reverse loss curves would show small distance changes. However, our findings in Figure 6 contradict this hypothesis and instead support assumption (d) in Figure 5: extreme cases primarily result in large horizontal shifts of the reverse loss dynamic, while the forward loss dynamic remains stable (simple vertical shift). In cases where forward loss substantially exceeds reverse loss (top-10% cases), we observe that reverse loss decreases rapidly, with the intersection point occurring very early in the sequence. Conversely, in cases where reverse loss is larger (bottom-10% cases), the intersection point is delayed until near the sequence end, with reverse loss consistently exceeding forward loss throughout most of the process.

The results shows that while the average loss difference is an aggregated metric, it effectively indicates different patterns in step-by-step loss dynamics. With our case studies in Table 1, we find the obvious text quality differences between the reverse-favoring cases and forward-favoring cases. This finding suggests that text quality is the key feature influencing the loss dynamics and the positions of intersection points. Our analysis also re-

Model & Strategy	Stem	Humanities	Social Science	Other	Average
Original Llama2-7B	35.84	50.60	50.46	48.10	45.29
CT w/ All SlimPajama-6B	36.15	46.74	49.03	46.63	43.85
CT w/ Random 1B	35.73	46.16	48.40	47.08	43.57
CT w/ PPL Lowest Ranked 1B	36.24	45.79	47.57	45.53	43.09
CT w/ $\mathcal{S}$ Lowest Ranked 1B	34.04	45.94	45.66	42.93	41.38
CT w/ $\mathcal{S}$ Highest Ranked 1B	**37.15**	**50.93**	**50.63**	**49.82**	**46.24**

Table 2: Results (Accuracy%) on the MMLU benchmark among different data selection strategies on LLaMA2-7b continued pre-training (CT). $\mathcal{S}$ is our proposed quality score simply computed by Forward Loss - Reverse Loss. More results with different backbones across various benchmarks are shown in Appendix C.
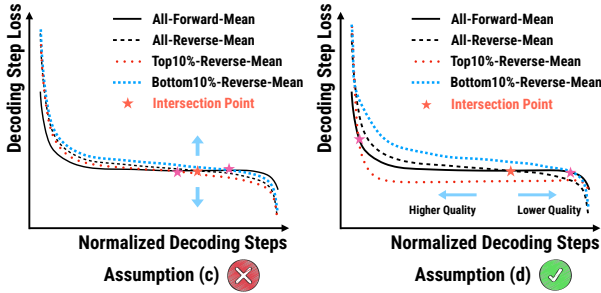

Figure 5: Assumptions on the step-by-step loss dynamics of selected texts with the Top-10% and Bottom-10% loss differences during decoding.
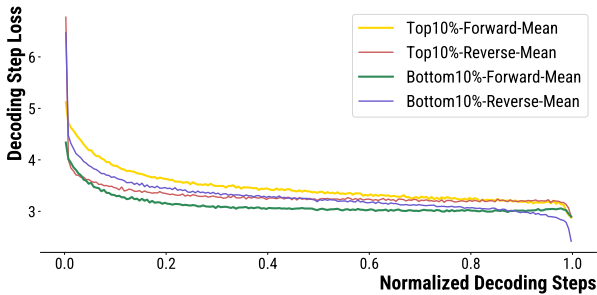

Figure 6: Empirical step-by-step loss dynamics of selected texts with the Top-10% and Bottom-10% loss differences during decoding.

veals a key insight: texts exhibiting early intersection points in their loss dynamics typically have higher loss differences and correspond to higher quality content. This relationship enables us to use the loss difference as a quality score for text quality assessment.

Application: *Texts favoring reverse modeling can improve original LLMs.*

As analyzed in RQ3, coherent and logical texts tend to have lower reverse losses compared to forward losses. Thus, given a training sample and a LLM model pre-trained from scratch with both forward and reverse training, we can define a simple quality score $\mathcal{S}$ using the loss difference $\mathcal{S} = $ Avg. Forward Loss - Avg. Reverse Loss, computed as $\mathcal{S} = \frac{1}{N}(\sum_{i=1}^{N} F_i - \sum_{i=1}^{N} N_i)$. Ac-

cording to our prior analysis, A higher $\mathcal{S}$ indicates that the text, which supports reverse modeling better, signifies a high-quality sample.

To further verify this assumption, we conduct continued pre-training on the publicly released Llama2-7B. Using the SlimPajama-6B (Soboleva et al., 2023) as training data, we select 1B tokens with the lowest and highest $\mathcal{S}$ scores, respectively. The model's performance is evaluated on MMLU (Hendrycks et al., 2020). We also compare this with the following data selection strategies: (1) **Random 1B**: randomly sample 1B tokens, (2) **Perplexity Lowest Ranked 1B**: select the 1B tokens with the lowest perplexity by Llama2-7B.

The results from Table 2 show that the quality of training data significantly affects the performance of LLMs. Our high-quality data selection strategy ($\mathcal{S}$ Highest Ranked) outperforms other baselines, achieving the highest accuracy across various tasks on MMLU. Since the overall text quality of the SlimPajama 6B dataset is inferior to the text quality used in the pretraining of Llama2-7B, using the full 6B dataset does not improve over the original Llama2-7B. This suggests that the presence of low-quality data in unfiltered training sets degrades performance, as evidenced by the significant performance decline with low-quality selection strategy ($\mathcal{S}$ Lowest Ranked). This experiment supports the hypothesis that texts more effectively modeled by reversing are of higher quality and more beneficial for LLMs in acquiring world knowledge.

5 Conclusions

In conclusion, our results demonstrate that LLMs can learn from both forward and reverse-modeling texts with comparable proficiency when trained from scratch. This study also highlights the potential benefits of incorporating training data that favors reverse modeling. Our findings underscore the importance of exploring diverse reverse modeling frameworks to enhance the capabilities of LLMs.

Limitations

While our study demonstrates promising results in training LLMs with reverse modeling, several limitations should be acknowledged to provide a comprehensive understanding of the findings and guide future research.

Firstly, the simulation of reverse modeling by simply reversing token sequences may not fully capture the complexity and nuances of true reverse thinking processes. This approach reduces reverse modeling to a syntactic level, potentially overlooking deeper semantic and contextual factors intrinsic to human reverse modeling.

Secondly, the evaluation metrics used in our study, such as performance on downstream benchmarks like MMLU, may not fully encompass the benefits or limitations of reverse modeling. These metrics primarily measure specific aspects of language understanding and reasoning, potentially overlooking other critical dimensions influenced by reverse modeling, such as creativity or problem-solving skills.

Lastly, our research does not address the potential computational and resource challenges associated with training LLMs on reverse texts. The increased complexity and processing demands could pose significant barriers to practical applications, particularly in resource-constrained environments.

In conclusion, while our findings offer valuable insights into the potential of reverse modeling in LLMs, addressing these limitations is crucial for advancing this line of research. Future studies could aim to develop more sophisticated methods for simulating reverse modeling, explore diverse and naturally occurring datasets, and consider a broader range of evaluation metrics to fully understand and harness the benefits of reverse modeling in LLMs.

Acknowledgments

References

Benjamin Bergen and Ting Ting Chan. 2005. Writing direction influences spatial cognition. In Proceedings of the Annual Meeting of the Cognitive Science Society, volume 27.

Lukas Berglund, Meg Tong, Maximilian Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. 2023. The reversal curse: Llms trained on "a is b" fail to learn "b is a". In Proceedings of the the 12th International Conference on Learning Representations (ICLR).

Justin Chih-Yao Chen, Zifeng Wang, Hamid Palangi, Rujun Han, Sayna Ebrahimi, Long Le, Vincent Perot, Swaroop Mishra, Mohit Bansal, Chen-Yu Lee, et al. 2024. Reverse thinking makes llms stronger reasoners. arXiv preprint arXiv:2411.19865.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2023. Palm: Scaling language modeling with pathways. Journal of Machine Learning Research (JMLR), 24(240):1–113.

Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. 2019. Boolq: Exploring the surprising difficulty of natural yes/no questions. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)).

Elizabeth Clark, Tal August, Sofia Serrano, Nikita Haduong, Suchin Gururangan, and Noah A Smith. 2021. All that's 'human'is not gold: Evaluating human evaluation of generated text. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP).

Joseph Cornelius, Oscar Lithgow-Serrano, Sandra Mitrović, Ljiljana Dolamic, and Fabio Rinaldi. 2024. Bust: Benchmark for the evaluation of detectors of llm-generated text. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL).

Derrick De Kerckhove and Charles J Lumsden. 2013. The alphabet and the brain: The lateralization of writing. Springer Science & Business Media.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783.

Olga Golovneva, Zeyuan Allen-Zhu, Jason Weston, and Sainbayar Sukhbaatar. 2024. Reverse training to nurse the reversal curse. In Proceedings of the 1st Conference on Language Modeling (COLM).

Jiatao Gu, Qi Liu, and Kyunghyun Cho. 2019. Insertion-based decoding with automatically inferred generation order. Transactions of the Association for Computational Linguistics (TACL), 7:661–676.

Qingyan Guo, Rui Wang, Junliang Guo, Xu Tan, Jiang Bian, and Yujiu Yang. 2024. Mitigating reversal curse via semantic-aware permutation training. In Findings of the Association for Computational Linguistics: ACL 2024.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. In Proceedings of the 8th International Conference on Learning Representations (ICLR).

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. arXiv:2310.06825.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

Shayne Longpre, Gregory Yauney, Emily Reif, Katherine Lee, Adam Roberts, Barret Zoph, Denny Zhou, Jason Wei, Kevin Robinson, David Mimno, et al. 2024. A pretrainer's guide to training data: Measuring the effects of data age, domain coverage, quality, & toxicity. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL).

Arindam Mitra, Hamed Khanpour, Corby Rosset, and Ahmed Awadallah. 2024. Orca-math: Unlocking the potential of slms in grade school math. arXiv preprint arXiv:2402.14830.

Anh Nguyen, Nikos Karampatziakis, and Weizhu Chen. 2024. Meet in the middle: A new pre-training paradigm. In Proceedings of the 38th Annual Conference on Neural Information Processing Systems (NeurIPS).

Vassilis Papadopoulos, Jérémie Wenger, and Clément Hongler. 2024. Arrows of time for large language models. arXiv preprint arXiv:2401.17505.

Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. 2023. The refinedweb dataset for falcon llm: outperforming curated corpora with web data, and web data only. In Proceedings of the 37th Annual Conference on Neural Information Processing Systems (NeurIPS).

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research (JMLR), 21(140):1–67.

Vasu Sharma, Karthik Padthe, Newsha Ardalani, Kushal Tirumala, Russell Howes, Hu Xu, Po-Yao Huang, Shang-Wen Li, Armen Aghajanyan, and Gargi Ghosh. 2024. Text quality-based pruning for efficient training of language models. arXiv preprint arXiv:2405.01582.

Daria Soboleva, Faisal Al-Khateeb, Robert Myers, Jacob R Steeves, Joel Hestness, and Nolan Dey. 2023. Slimpajama: A 627b token cleaned and deduplicated version of redpajama.

Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, Aakanksha Chowdhery, Quoc Le, Ed Chi, Denny Zhou, et al. 2023. Challenging big-bench tasks and whether chain-of-thought can solve them. In Findings of the Association for Computational Linguistics: ACL 2023.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv:2307.09288.

Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. 2020. Ccnet: Extracting high quality monolingual datasets from web crawl data. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC).

Alexander Wettig, Aatmik Gupta, Saumya Malik, and Danqi Chen. 2024. Qurating: Selecting high-quality data for training language models. In Proceedings of the 41st International Conference on Machine Learning (ICML).

Lijun Wu, Xu Tan, Di He, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018. Beyond error propagation in neural machine translation: Characteristics of language also matter. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Sang Michael Xie, Shibani Santurkar, Tengyu Ma, and Percy S Liang. 2023. Data selection for language models via importance resampling. In Proceedings of the 37th Annual Conference on Neural Information Processing Systems (NeurIPS).

Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. 2024. Tinyllama: An open-source small language model. arXiv preprint arXiv:2401.02385.

Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. 2024. Agieval: A human-centric benchmark for evaluating foundation models. In Findings of the Association for Computational Linguistics: NAACL 2024.

Long Zhou, Jiajun Zhang, and Chengqing Zong. 2019. Synchronous bidirectional neural machine translation. Transactions of the Association for Computational Linguistics (TACL), 7:91–105.

Hanlin Zhu, Baihe Huang, Shaolun Zhang, Michael Jordan, Jiantao Jiao, Yuandong Tian, and Stuart Russell. 2024. Towards a theoretical understanding of the'reversal curse'via training dynamics. In Proceedings of the 38th Annual Conference on Neural Information Processing Systems (NeurIPS).

A Implementation Details

In our initial experiments, we explore the effects of a higher learning rate ($2e^{-4}$) and extend the training epochs (2 epochs) in a continued training setting. This exploration is based on the assumption that reverse modeling might need more epochs or a higher learning rate than forward modeling to overcome the pre-trained directional bias. While this increases the convergence speed, the final loss is nearly the same as when using a lower learning rate or a single training epoch.

Thus, to ensure consistency in our comparisons, we fix the learning rate as $5e^{-5}$ and set the batch size to 48 (with each batch consisting of 48 paragraphs). Following established practices in LLM training (Touvron et al., 2023; Chowdhery et al., 2023), we train for one epoch in all experiments. The number of training steps depends on the size of the data set and the batch size. For instance, training on a dataset with 1 billion tokens requires approximately 2, 400 steps with our hyperparameter settings.

We use Llama2-7B (Touvron et al., 2023) as the LLM backbone for Research Questions 1–3 (Section 4). All experiments are conducted using 8 NVIDIA A100-SXM-80GB GPUs, and the application experiments in Section 4 also use the Llama2-7B model.

B Multilingual Experimental Results

We show the pre-training losses for both continued and from-scratch training across additional languages, including German, Korean and Arabic, in Figure 7. Note that Arabic texts tokenized by Llama2 tokenizer have the same orientation as English, with tokens from the first logical sentence of a paragraph positioned on the left rather than the right. Consistent with our findings in RQ1, Section 4, the forward loss during continued pre-training remains lower than the reverse loss. However, in the from-scratch setting, the loss curves for both directions converge similarly. These results further confirm that LLMs can effectively learn to handle both forward and reverse inputs with comparable proficiency when trained from scratch, regardless of languages.

We randomly sample additional multilingual cases (German, Korean, and Arabic), as shown in Tables 4-7. Across all four languages, we observe a consistent pattern: Texts favoring reverse modeling tend to exhibit clear logical structures, while those

Model & Strategy	MMLU	AGIEval	BBH	BoolQ
Llama2-7B				
Random 1B	43.57	26.53	42.33	74.86
Lowest Ranked 1B	41.38	25.56	38.26	73.96
Highest Ranked 1B	**46.24**	**27.07**	**43.79**	**75.44**
Mistral-7B				
Random 1B	35.45	40.97	43.45	77.34
Lowest Ranked 1B	34.99	38.85	42.17	75.29
Highest Ranked 1B	**36.66**	**42.86**	**44.98**	**78.56**
Llama3-8B				
Random 1B	58.94	-	-	-
Lowest Ranked 1B	58.54	-	-	-
Highest Ranked 1B	**59.49**	-	-	-

Table 3: Experimental results using three different LLM backbones on the MMLU, AGIEval, BBH, and BoolQ benchmarks. However, we exclude Llama3-8B's results on AGIEval, BBH, and BoolQ, as the evaluation sets for these benchmarks are found to overlap significantly with its training data (contaminated rates: 98%, 95%, and 96%, respectively) (Dubey et al., 2024).

favoring forward modeling rely more on repetitive formatting to emphasize their sequential flow.

C More Results with Different LLM Backbones on Different Benchmarks

Besides the MMLU (Hendrycks et al., 2020) benchmark, we also evaluate our proposed data selection application on three benchmarks, *i.e.*, AGIEval (Zhong et al., 2024), BBH (Suzgun et al., 2023) and BoolQ (Clark et al., 2019), using different LLM backbones including Llama2-7b (Touvron et al., 2023), Mistral-7b (Jiang et al., 2023) and Llama3-8b (Dubey et al., 2024).

The experimental results are shown in Table 3. Our high-quality data selection strategy (Highest Ranked) consistently outperforms other approaches across various benchmarks, regardless of the LLM backbone used. These results support our hypothesis that texts better modeled by reverse prediction yield higher quality data, which in turn enhances the LLMs' ability to acquire world knowledge. Notably, the ranked score, $\mathcal{S}$ = Forward Loss - Reverse Loss, is computed and fixed using the Llama2-7b model throughout the experiments. This highlights the strong generalization capability of our method, as the high-quality data selected by one LLM can be effectively transferred to another.
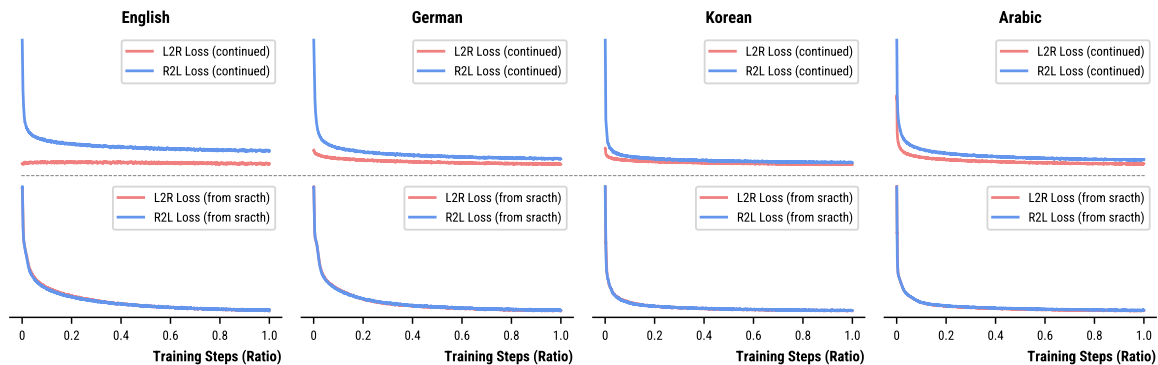
Figure 7: Pre-training losses for both continued and from-scratch training settings in four additional languages. The patterns are consistent with the results observed in English.

System Prompt:

You are an expert in text quality checking. You need to score a given text used for large language model training from 1 to 10 according to the following factors:

1. *Grammar*: The spelling and grammar of the text, punctuation/formatting issues.
2. *Level of detail*: Is the text simple or does it go more in-depth?
3. *Genre*: If the text is the genre/domain/style/formality that the reader expects, adheres to style norms.
4. *Repetition*: Words/phrases/content repeated itself.
5. *Factuality*: The accuracy of the text, whether it describes things that are "true."
6. *Consistency*: How the text relates to the context and other pieces of the text.
7. *Common Sense*: Whether the text "makes sense" within the world that it is written.
8. *Coherence*: The structure and coherence of the text. Order issues go here.
9. *Writer intent and expression*: Speculating about writer's intent or capabilities.
10. [Most Important Factor] *Quality for LLM Training*: this text will be used for LLM Training by causal language modeling.

Please remember to give score strictly, score to differentiate the samples, and prevent to give most of the cases similar score.
Output in one-line JSON format: {"score": "<score>", {"reason": "<reason>"}

Figure 8: The prompt used for text qualitative evaluation using GPT-4 API. The 1-9 factors follows the designed evaluation labels in (Clark et al., 2021), and we add an extra "Quality for LLM Training" into the evaluation factors.

D Qualitative Analysis

Many open-source LLMs use data classifiers for selection and cleaning, but the specifics of their training processes are often proprietary and not fully detailed in technical reports (Touvron et al., 2023; Jiang et al., 2023; Dubey et al., 2024). Thus, we expand our experiments using the GPT-4 API [3] to directly evaluate the quality of texts from the Lowest Ranked 1B and Highest Ranked 1B texts, in line with previous studies (Clark et al., 2021; Cornelius et al., 2024). We randomly select $1,000$ samples from each dataset and apply a predefined prompt (shown in Figure 8) to assess each sample. The GPT-4 API assigns a quality score ranging from 1 to 10, based on criteria for high-quality

text defined in (Clark et al., 2021), along with an additional criterion for "quality for LLM training".

The results show that the Highest Ranked 1B dataset achieves an average score of 6.7, while the Lowest Ranked 1B and Random 1B datasets scored 4.9 and 6.15, respectively. These findings further suggest that texts favoring reverse modeling are of higher quality and more suitable for LLM training.

E Smaller LLM as Backbone Model

We also conduct experiments on the TinyLlama-1.1B model (Zhang et al., 2024) under the same protocols used in Section 4 4. Figure 9 shows that the loss trend of TinyLlama-1.1B is similar to that of the larger Llama2-7B model. It indicates that even with smaller models, both forward and reverse modeling can be effectively handled in the from-scratch training setting. Next, we recalculate the

---

[3]https://platform.openai.com/docs/api-reference/
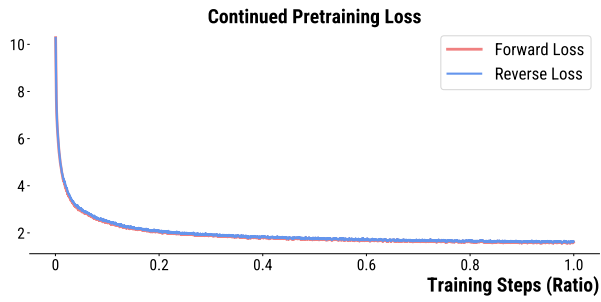
Continued Pretraining Loss

Figure 9: Pre-training loss in the from-scratch setting using English data on TinyLlama-1.1B. The forward and reverse losses are nearly identical, aligning with findings from larger LLM models.

quality score $\mathcal{S}$ = Forward Loss - Reverse Loss for each case of the SlimPajama (Soboleva et al., 2023) dataset and select the lowest-ranked 1B and highest-ranked 1B data based on the recalculated scores. Notably, the overlap ratios of the selected cases are 91.27% and 94.58% when compared to Llama2-7B. This demonstrates that our method is model-agnostic, enabling smaller models to efficiently identify high-quality data for training larger models in practice.

Language	Texts Favoring Reverse (Low Reverse Loss)	Texts Favoring Forward (Low Forward Loss)
English	Have you ever wondered why cultures in the hottest locations on earth eat hot and spicy foods? Why is it that people in Central and South America, India, Africa, Southeast Asia and the Caribbean eat foods flavored with hot chile peppers and spices that make you sweat? There is a reason, and it's actually pretty smart when you think about it — spicy foods make you sweat, which in turn helps you cool down faster. It's as simple as that! Though you may be inclined to cool down with a tall glass of iced tea, ice cream or watermelon on a sweltering summer's day, the effect isn't lasting. After a while you're back to where you started — hot and bothered. That's because your internal temperature is cooled too rapidly, and your body ends up compensating by raising your temperature. As a result, you feel hotter. Eating spicy foods works differently — it raises your internal temperature to match the temperature outside. Your blood circulation increases, you start sweating and once your moisture has evaporated, you've cooled off. Scientists call the phenomenon "gustatory facial sweating," because indeed you usually start sweating in the face first. Even though eating spicy foods on a hot day isn't the most pleasant for many people, it may be worth doing because after sweating it out you do actually cool down. What do you think: is it worth it?	GET $2000 CASHBACK!!!! Nissan Qashqai combines stunning looks, efficient aerodynamics, and advanced technology to help you enjoy enlightened driving at its best. And thanks to Nissan Intelligent Mobility, you'll feel more confident and connected than ever. Loaded with the state of the art features including: -5 Star ANCAP Safety Rating -Forward-Collision Warning -Flat-Bottom Steering Wheel -Black Leather-accented Seat and Steering Wheel Trim -Individually Heated Front Seats -Dual Zone Climate Control -7?809D Touch Screen Display -Satellite Navigation -Digital (DAB) Radio -Intelligent Around-View Cameras -Blind Spot Alert -Rear Cross-Traffic Alert -Rear Privacy Glass -Fog Lights -Roof Rails -18?809D alloy wheels -LED Daytime Running Lights and Taillights -ISOFIX Child Restraint anchorage -Vehicle Dynamic Control -Cruise control with digital speedometer -Bluetooth hands free system with audio streaming -6 speaker sound system -AUX/iPod connectivity -Power windows -Power mirrors and much more! Located on Road in , close to public transport and freeways, and only a 25 minute drive from the CBD, we have been selling and servicing Nissan vehicles across Melbourne for over 25 years.
	The kinetic equations for clean superconductors ($I \gg \xi$) are derived. Expanding the equations for the time dependent Green functions in the quasiclassical parameter, the new contributions are found which contain the derivatives of the distribution functions with respect to the quasiparticle momentum. The transition from the ultra-clean case (no relaxation) to a relaxation-dominated behavior, for which the kinetic equations coincide with the usual quasiclassical approximation, occurs for the relaxation time of the order of $\hbar E_F/\Delta 2$. The kinetic equations can be used for various dynamic processes in superconductors including the flux-flow Hall effect. The derived equations, after necessary modifications for the p-wave pairing, are especially suitable for nonstationary problems in the theory of superfluidity of 3He.	Our Cartridges for Lexmark X2250 are great value with super fast delivery! Cartridges for Lexmark X2250 are among our thousands online products. With our huge range and simple website, it is easy to find all the cartridges you need for any other printers you may have. Together with our most competitive prices, we are sure to be your one-stop online store! Cartridges for Lexmark X2250 are covered by a 60 days warranty. If the product you received is faulty, please contact us to organise a replacement or refund. Please refer to our Warranty Return. When will my Cartridges for Lexmark X2250 be delivered? In most cases you will receive your Cartridges for Lexmark X2250 the next working day, or within 3 days if outside the next day express post network. It might takes up to 6 days for some remote areas.
	How is this club different from standard Toastmasters Clubs? As an Advanced Toastmasters Club, Professional Speakers Frankfurt is only open to experienced members who have completed the Toastmasters Competent Communicator level or are advanced speakers with proven experience outside the Toastmasters world. Therefore, we can focus on more advanced issues. Rather than having the majority of speeches from the Competent Communicator manual we focus on advanced projects or practice speeches outside of Toastmasters manuals. We prepare members for speech competitions or help someone with an upcoming professional or other important speech. Instead of having only one evaluator, all attendees will get the chance to provide feedback. Depending on the objectives of the speaker, the group might be divided into task forces to keep an eye on particular aspects and debrief the speaker afterward. We also use video recording to provide in-depth analysis of a speaker's performance. Sounds boring? In fact, it isn't. For a speech to become great, it has to go through multiple iterations. In our club, we give members the opportunity to repeat a speech and incorporate the feedback they've received. We hold regular advanced workshops run by members or outside experts on specific speech-related topics. We encourage our members to participate in Toastmasters speech contests and dedicate special time to prepare candidates. We will develop a peer coaching system through which members continuously coach each other. We will set up a Speakers Bureau, and members will be able to present and promote themselves and as speakers on the Club website and through the Club's online channels.	Rent a Dumpster in Oswego Now! Simply give us a call and we will answer any questions you may have about the Oswego dumpster rental procedure, allowed materials, size you may need, etc. Our roll off containers can usually be delivered in Oswego the day after you place your order. Make sure to call us early to ensure timely delivery. Whether or not you require a long-term or roll-off dumpster is dependent upon the type of job and service you need. Long-Term dumpster service is for ongoing demands that last more than simply a few days. This includes matters like day-to-day waste and recycling needs. Temporary service is precisely what the name suggests; a one time need for project-special waste removal. Temporary roll off dumpsters are delivered on a truck and are rolled off where they'll be utilized. These are typically larger containers that may manage all the waste that comes with that specific job. Long-Term dumpsters are generally smaller containers because they're emptied on a regular basis and so don't need to hold as much at one time. Should you request a permanent dumpster, some firms require at least a one-year service agreement for this dumpster. Rolloff dumpsters only require a rental fee for the time that you keep the dumpster on the job. If you want to rent a dumpster in Oswego, you will find that costs vary significantly from state to state and city to city. One means to get genuine estimates for the service you need would be to telephone a local dumpster company and ask regarding their costs. You can also request a quote online on some sites. These sites may also contain full online service that is constantly open. On these sites, you can choose, schedule and pay for your service whenever it's convenient for you. Variables which affect the price of the container contain landfill fees (higher in some areas than others) as well as the size of the container you opt for. You also need to consider transportation costs as well as the kind of debris you will be placing into your container.
	Complete replacement of factory floor automation systems program for the largest auto plant in North America. It was a body assembly plant, a paint plant and a final assembly plant. The size of the plant was being doubled to accommodate a new model. With less than three months to go, the launch was in jeopardy because the systems were not ready for installation. Failure to install would delay the new model six months. An unsuccessful installation would shut the plant down.	With SoundcloudToMp3 you can convert and download music in High Quality MP3 format. Download tons of music from Soundcloud with our Soundcloud Downloader and listen to them from anywhere by storing them on your iPod, computer or phone using our ultra fast downloading service. SoundCloud is audio distribution site, where users can record, upload and promote their sound tracks. SoundCloud allows you to listen as many tracks you can but it does not allow sound track download. Enter the Soundcloud URL that you wish to convert & Download. Click "Convert it" to start the conversion process. Click "Download Mp3" to download the file. Once complete you will have final download link for converted sound. Highly Secure and high speed. Mp3 Converter supports a wide variety of modern browsers and devices.

Table 4: Case study in English. The texts favoring reverse are typically high-quality and well-suited for LLM training. In contrast, those favoring forward modeling often exhibit repetition and occasional lapses in logic and coherence, which can negatively impact LLM training.

Language	Texts Favoring Reverse (Low Reverse Loss)	Translation	Texts Favoring Forward (Low Forward Loss)	Translation
	In aller Munde, in aller Ohren – an Jonas Kaufmann kommt man derzeit nicht vorbei. Startenor, Herzensbrecher, ein echtes Münchner Kindl noch dazu, hat sich Kaufmann in die internationale erste Riege gesungen. „Seine Intensität und seine Eleganz, die Geschmeidigkeit seiner Stimme und seiner Körpersprache, kombiniert mit seiner Musikalität und seinem strahlenden Aussehen, machen ihn zum Inbegriff des Opernstars im 21. Jahrhundert", schwärmte der Herausgeber der Opera News. Und so wird Jonas Kaufmann seit geraumer Zeit weltweit gefeiert – nicht nur an den größten Opernhäusern, sondern auch als Protagonist in Gustav Mahlers „Lied von der Erde", als Interpret von Hugo Wolfs „Italienischem Liederbuch" oder als leidenschaftlicher Tenor, wenn er in einer Hommage an die unsterbliche Musik Italiens ihren Evergreens eine besondere Magie verleiht. ...	On everyone's lips, in everyone's ears – it's impossible to overlook Jonas Kaufmann at the moment. Star tenor, heartthrob, and a true Munich native, Kaufmann has sung his way into the international top ranks. 'His intensity and elegance, the smoothness of his voice and body language, combined with his musicality and his radiant appearance, make him the epitome of the 21st-century opera star,' enthused the editor of Opera News. And so, Jonas Kaufmann has been celebrated worldwide for quite some time – not only at the greatest opera houses, but also as the lead in Gustav Mahler's "Das Lied von der Erde", as an interpreter of Hugo Wolf's "Italian Songbook", or as a passionate tenor when he lends a special magic to Italian evergreens in a tribute to the immortal music of Italy. ...	... Urlaubsangebote für Yaroslavl Spielen Sie mit dem Gedanken, eine Reise nach Yaroslavl zu buchen? Ob Sie einen Romantikurlaub, eine Familienreise oder ein All-Inclusive-Paket planen, die Pauschalreisen nach Yaroslavl auf TripAdvisor machen die Reiseplanung einfach und erschwinglich. Vergleichen Sie Hotel- und Flugpreise für Yaroslavl und finden Sie so auf TripAdvisor die perfekte Pauschalreise nach Yaroslavl. Reisende wie Sie haben 7.983 Bewertungen geschrieben und 10.284 authentische Fotos für Hotels in Yaroslavl gepostet. Buchen Sie Ihren Urlaub in Yaroslavl noch heute! Familienfreundliche Hotels in Yaroslavl „Gute Lage, ein Park und Kotorosl Ufer fußläufig gut erreichbar. Zimmer sind sauber und werden immer gut aufgeräumt. Ein sehr bequemes Bett, das man sehr selten findet. Auch einen sehr guten und ...	... Holiday Offers for Yaroslavl Are you thinking about booking a trip to Yaroslavl? Whether you are planning a romantic getaway, a family trip, or an all-inclusive package, the vacation packages to Yaroslavl on TripAdvisor make planning your trip easy and affordable. Compare hotel and flight prices for Yaroslavl, and find the perfect package on TripAdvisor. Travelers like you have written 7,983 reviews and posted 10,284 authentic photos of hotels in Yaroslavl. Book your vacation to Yaroslavl today! Family-Friendly Hotels in Yaroslavl "Good location, with a park and the Kotorosl Riverbank within walking distance. The rooms are clean and always well-maintained. A very comfortable bed, which is hard to find. Also, a very good and..." ...
German	... Elisabeth von Luxemburg wurde 1422 13jährig mit dem 25 Jahre alten Thronanwärter Albrecht V. verheiratet (verlobt waren sie bereits seit ihrem 2. Lebensjahr). Nach den ersten zehn Jahren Ehe bekam sie ihr erstes von vier Kindern; fünf Jahre später wurde ihr Gemahl durch den Tod seines Vaters römisch-deutscher König sowie König von Ungarn, Kroatien und Böhmen. Elisabeth war im fünften Monat mit dem vierten Kind schwanger, als er 1439 während eines Feldzuges gegen die in Ungarn einfallenden Türken an der Ruhr verstarb. Entgegen dem politischen Drängen des Adels, den 15jährigen polnischen König Wladislaw III. zu heiraten – weil ein männlicher König gleich welchen Alters und Charakters für das Land im Krieg gegen die Türken „sicherer" sei –, ergriff sie selbst die Regentschaft, um so bald als möglich ihren Sohn Ladislaus Postumus zum König zu machen. Bevor der Adel Wladislaw per Königswahl vor ihren Sohn setzen konnte, bemächtigte sich Elisabeth der Stephanskrone, die als heilig betrachtet wurde und deren Besitz den König von Ungarn legitimierte. Hierfür sandte sie ihre Kammerfrau Helene Kottannerin in die Plintenburg, aus der die Kottannerin die Insignie erfolgreich entführte und mit einer Schlittenfahrt über die gefrorene Donau (es war Februar) zu ihrer Königin brachte. Die Kottannerin schrieb darüber später in ihren Memoiren „Denkwürdigkeiten". Elisabeth krönte ihren Sohn zum König von Ungarn, Kroatien und Böhmen und behielt die Stephanskrone auch, nachdem sie sie eigentlich hatte zurückgeben sollen, durch einen Betrug in ihrem Besitz. ...	... Elisabeth of Luxembourg was married to the 25-year-old heir to the throne, Albert V, in 1422 at the age of 13 (they had been betrothed since she was 2 years old). After the first ten years of marriage, she gave birth to the first of their four children. Five years later, upon the death of his father, her husband became King of the Romans (Holy Roman Emperor-elect), as well as King of Hungary, Croatia, and Bohemia. Elisabeth was five months pregnant with their fourth child when her husband died in 1439 during a military campaign against the Turks, who were invading Hungary. Despite political pressure from the nobility to marry the 15-year-old Polish king Wladislaw III—because having a male king, regardless of his age or character, was seen as "safer" for the country in the war against the Turks—she took on the regency herself. Her goal was to secure the throne for her son, Ladislaus Postumus, as quickly as possible. Before the nobility could elect Wladyslaw as king over her son, Elisabeth took possession of the Holy Crown of Hungary, which was regarded as sacred and essential for legitimizing the king of Hungary. To achieve this, she sent her chambermaid, Helene Kottanner, to Visegrád (Plintenburg), from where Kottanner successfully stole the crown and delivered it to her queen by sled across the frozen Danube (it was February). Kottanner later recounted this event in her memoirs, Memorabilia. Elisabeth crowned her son as King of Hungary, Croatia, and Bohemia. Even after she was supposed to return the Holy Crown, she kept it in her possession through deceit. ...	... Entdecken Sie, wie viel eine Busfahrt von Mundo Novo nach Maracaju kostet. Verwenden Sie unsere Filter und Sortierfunktionen, um die billigsten Bus-Tickets von Mundo Novo nach Maracaju, oder Luxus-Fernbusse zu finden. Busse, die von Mundo Novo nach Maracaju fahren, starten von der Station Terminal Rodoviaria Mundo Novo. Ein Bus nach Maracaju wird Sie an der Station Maracaju Onibus absetzen. Streckenplan Mundo Novo nach Maracaju Wenn Sie im Ausland sind, sollten Sie auch etwas von der Landessprache lernen. Auf Ihrer Busreise von Mundo Novo nach Maracaju könnte das in einer misslichen Lage sehr nützlich sein und die einheimische Bevölkerung wird sich bestimmt über Ihre Anstrengungen, eine neue Sprache zu lernen, freuen. Freuen Sie sich bei Ihrer Busreise von Mundo Novo nach Maracaju auf einen wahren Augenschmaus mit wunderschönen Naturlandschaften und eindrucksvollen Sehenswürdigkeiten auf vielen Kilometern. Busse haben von alle motorisierten Fortbewegungsmitteln den geringsten CO2-Ausstoss. Ein Fernbus von Mundo Novo nach Maracaju wird im Vergleich zu einem Zug nur halb so viel CO2 ausstoßen, und die Bilanz sieht im Vergleich zum Auto oder einem Flugzeug sogar noch wesentlich besser aus. Erstellen Sie einen Soundtrack für Ihr eigenes Leben, indem Sie eine personalisierte Playlist für die Busreise erstellen. Kann es einen besseren Begleiter für Ihre Busfahrt von Mundo Novo nach Maracaju geben als Ihre Musik? ...	... Discover how much a bus ride from Mundo Novo to Maracaju costs. Use our filters and sorting features to find the cheapest bus tickets from Mundo Novo to Maracaju, or opt for luxury coaches. Buses traveling from Mundo Novo to Maracaju depart from the Terminal Rodoviaria Mundo Novo station. A bus to Maracaju will drop you off at the Maracaju Onibus station. Route Plan: Mundo Novo to Maracaju If you are traveling abroad, it's a good idea to learn some of the local language. On your bus journey from Mundo Novo to Maracaju, this could be very helpful in an emergency, and the locals will surely appreciate your efforts to learn a new language. Look forward to a visual feast on your bus journey from Mundo Novo to Maracaju, with stunning natural landscapes and impressive sights stretching over many kilometers. Of all motorized modes of transportation, buses have the lowest CO2 emissions. A coach from Mundo Novo to Maracaju will emit only half as much CO2 as a train, and the environmental impact compared to a car or airplane is even better. Create a soundtrack for your life by making a personalized playlist for your bus journey. Could there be a better travel companion for your trip from Mundo Novo to Maracaju than your music? ...

Table 5: Case study in German. Included are both the original German texts and their English translations.

Language	Texts Favoring Reverse (Low Reverse Loss)	Translation	Texts Favoring Forward (Low Forward Loss)	Translation
Korean	"미국 동영상 서비스 시장, 최종 승자는 누구?' - B2B IT 전문가 진행 생방송토크 웨비나 전 세계에서 인터넷 동영상 서비스(Over The Top, OTT) 경쟁이 한창이다. 글로벌 온라인 동영상 스트리밍 서비스의 선두주자 넷플릭스, 아마존닷컴의 인터넷 주문형 동영상 서비스 아마존 비디오, 동영상 공유 사이트 유튜브 등 각자의 서비스를 내세우며 피 터지는 경쟁을 하고 있다. 중심지는 아무래도 미국이다. 글로벌 IT기업의 집결지인 미국 무대를 먼저 사로잡아야 전 세계 고객들을 사로잡을 수 있다는 생각으로 오리지널 콘텐츠 개발 등 각종 공격적 마케팅 전략을 쏟아내고 있다. 콘텐츠 개발을 위한 투자 예산도 어마어마하다. 지난 4월7일 <비즈니스인사이더> 보도에 따르면, 아마존이 2017년 동영상 서비스 강화를 위해 투입할 예산이 45억달러, 우리 돈 5조1천억원 규모라는 JP모건 애널리스트들의 분석이 나왔다. 브라이언 올사브스키 아마존 CFO 역시 "아마존 비디오에 대한 투자를 두 배 가까이 늘릴 것"이라고 말한 바 있다. 넷플릭스도 만만치 않다. 넷플릭스는 지난해 말, 2017년 서비스 강화를 위해 50억달러, 우리돈 5조7천억원 규모를 투입할 예정이라고 말했다. 두 회사의 투자 규모만 합쳐도 우리돈 12조원 정도 예산이니 가히 엄청나다고 할 수 있다. (자료=컴스코어) 미국 인터넷 시장조사 연구기업 컴스코어가 OTT 서비스 시장에 대한 조사 보고서를 4월10일 내놓았다. 컴스코어에 따르면, 2016년 12월을 기준으로 미국 내에 인터넷 연결망을 가진 가구 중 53%인 약 4900만 가구가 인터넷 동영상 서비스에 가입했다고 한다. 단순히 가입 규모에 그치지 않는다. 이들의 전체 평균 시청 시간은 월 평균 19일, 일 평균 2.2시간이다. 현재 미국인들의 하루 평균 TV 시청 시간은 4시간 수준이다. 케이블 위성 방송으로만 TV를 시청하던 전통적인 시청 패턴이 완전히 변화하고 있음을 알 수 있다. ...	"US video service market, who will be the final winner? - Live talk webinar hosted by B2B IT experts Competition in Internet video services (Over The Top, OTT) is in full swing around the world. Netflix, the leader in global online video streaming services, Amazon.com's Internet video-on-demand service Amazon Video, and video sharing site YouTube are competing fiercely by offering their own services. The center is obviously the United States. They are pouring out various aggressive marketing strategies, including the development of original content, with the belief that they can captivate customers around the world only by capturing the American stage, the gathering place of global IT companies, first. The investment budget for content development is also enormous. According to a report by <Business Insider> on April 7, JP Morgan analysts analyzed that Amazon's budget to invest in strengthening video services in 2017 is $4.5 billion, or 5.1 trillion won. Amazon CFO Brian Olsavsky also said, "We will nearly double our investment in Amazon Video." Netflix is no slouch either. Netflix said at the end of last year that it plans to invest $5 billion, or 5.7 trillion won, to strengthen its services in 2017. The combined investment size of the two companies alone amounts to a budget of approximately 12 trillion won, which can be said to be truly enormous. (Data = ComScore) ComScore, an American internet market research company, released a research report on the OTT service market on April 10. According to ComScore, as of December 2016, approximately 49 million households, or 53% of households with an Internet connection in the United States, had subscribed to Internet video services. It's not just about the size of subscriptions. Their overall average viewing time is an average of 19 days per month and 2.2 hours per day. Currently, the average amount of time Americans watch TV per day is around 4 hours. It can be seen that the traditional viewing pattern of watching TV only through cable and satellite broadcasting is completely changing. ...	"[37% 세일] Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 쿠폰 코드 인기 쿠폰, Jul 2020 - iVoicesoft 인기 쿠폰 › Cdkeys 쿠폰 코드 2020 › Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 쿠폰의 할인 코드 37% 세일, 여름 제공 간단히 버튼을 클릭하십시오 [할인 된 가격으로 즉시 구매] 쿠폰을 사용하려면 37% 할인 코드. 쿠폰 코드가 포함되었습니다. 결제시 코드를 입력하십시오. 특별승진의 (16.42Φ")16.42 절약 여름은 위대하신 Cdkeys 제공 받기에 완벽하시기입니다. 2020 년 여름 제공 위해 지금 청구하십시오. 현재 거래: 37% 할인 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US). Cdkeys 에서 원하는 것을 가져올 수있는 최고의 기회. 제한된 시간 동안 만. 결제시 코드를 입력하십시오. 최고의 세일즈 프로모션 사용하여 매력적인 가격으로 훌륭한 제품을 찾으십시오. 37% 할인 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US), 16.42 절약. 쇼핑하려면 클릭하세요. 제한된 시간 동안 만. Star Wars Battlefront II 2 - Celebration Edition Xbox One (US)에 대하여 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US)소개 Get 37% OFF of Star Wars Battlefront II 2 - Celebration Edition Xbox One (US), a 위대하 in 여름 제공 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 쿠폰 코드. Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 위대하 여름 제공 37% 쿠폰 코드. 왜 우리의 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 쿠폰 코드를 적용해야 합니까? 간단 해! 최신 Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 프로모션 코드를 수집하여 제공했습니다, 가장 큰 할인으로. 또한 모든 Cdkeys 제품에 대한 최고의 절감 효과를 제공합니다. Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) 할인 코드에 대한 의견"	"[37% Sale] Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Coupon Code Popular Coupons, Jul 2020 - iVoicesoft Popular Coupons › Cdkeys Coupon Codes 2020 › Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Coupon Discount Discount Code 37% Sale, Summer Offer Simply click the button [Buy Instantly at Discounted Price] to use coupon 37% discount code. Coupon code included. Enter code at checkout. Special Promotion (Save $16.42)Save $16.42 Summer is the perfect time to get great Cdkeys offers. Claim now for summer 2020 offer. Current deal: 37% off Star Wars Battlefront II 2 - Celebration Edition Xbox One (US). Best chance to get what you want from Cdkeys. Only for a limited time. Enter code at checkout. Cdkeys Coupon Code: Find great products at attractive prices using our best sales promotions. 37% off Star Wars Battlefront II 2 - Celebration Edition Xbox One (US), saving 16.42. Click to shop. For a limited time only. About Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Star Wars Battlefront II 2 - Celebration Edition Xbox One (US)Introduction Get 37% OFF of Star Wars Battlefront II 2 - Celebration Edition Xbox One (US), a great offer in summer Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Coupon Code. Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Greatest Summer Offer 37% Coupon Code. Why apply your Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) coupon code? It's simple! We have collected and provided you with the latest Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) promo codes, with the biggest discounts. We also offer the best savings on all Cdkeys products. Opinions on Star Wars Battlefront II 2 - Celebration Edition Xbox One (US) Discount Code"

Table 6: Case study in Korean. Included are both the original Korean texts and their English translations.

Language	Texts Favoring Reverse (Low Reverse Loss)	Translation	Texts Favoring Forward (Low Forward Loss)	Translation
Arabic	آبواسطة سبادتر آخر تحديث الثلاثاء، ١٠ مارس ٢٠١٥ صرح الممثل شون بين بأنه لن يعتذر إطلاقا عن تعليقه الساخر في حفل جوائز الأوسكار الذي واجه هجوما بسببه و ذلك عندما قال عن المخرج المكسيكي أليخاندرو غونزاليس إيناريتو من الذي أعطى لهذا اللعين الغرين كارد و ذلك خلال تقديمه لجائزة أفضل فيلم سينمائي والتي فاز بها المكسيكي عن فيلمردمن الذان عملا سويا في فيلم ٢١ جرمس. شون قال خلال الحملة الترويجية لفيلمه الجديد جنمن بأنه يتعجب من الإنتقادات المتكررة رغم تأكيده على أن أليخاندو هو إحدى أهم صانعي الأفلام في أمريكا و هذا التعليق لم يحمل أي نوايا عنصرية وغايته فقط هي إثارة الضحكات. يذكر أن الثنائي بين وغونزاليس قد عملا معا من قبل في فيلم ٢١ جرمس و قد صرح المخرج المكسيكي مباشرة عقب حفل الأوسكار أن هناك علاقة صداقة وطيدة تجمعه مع الممثل لذا هو يتقبل هذا التعليق بصدر رحب ويعتبرها مزحة مضحكة جدا. ردمنأليخاندرو غونزاليس أناريتوشون بينْ	"By sbo-editor last updated Tuesday, March 10, 2015 Actor Sean Penn said that he will never apologize for his sarcastic comment at the Academy Awards, for which he faced attacks, when he said about Mexican director Alejandro Gonzalez Inarritu, "Who gave this motherfucker the green card?" during his presentation of the award for Best Motion Picture, which the Mexican won. About the movie Birdman, they worked together in the movie 21 Grams. Sean said during the promotional campaign for his new movie, Gunman, that he is surprised by the repeated criticism, despite his assertion that Alejando is one of the most important filmmakers in America. This comment did not carry any racist intentions and was only intended to provoke laughter. It is noteworthy that the duo, Ben and Gonzalez, had previously worked together in the film 21 Grams, and the Mexican director stated immediately after the Oscar ceremony that there is a close friendship between him and the actor, so he accepts this comment with open arms and considers it a very funny joke. Birdman Alejandro Gonzalez Anaritoshon Bea"	إشتر عطر شانيل ألور سنشوال للنساء أو دي برفيوم – ١٠٠ مل أونلاين ـ إيزي كليك شانيل الور سنشوال عطر نسائي راقي، رائحته شرقية زهرية جميلة وساحرة، عبيره فواح وقوي الثبات ويضيف لمسات أنيقة ملفتة للأنظار ولا تقاوم وتدوم لفترات طويلة. انظر الوصف لمزيد من المعلومات شانيل الور سنشوال عطر نسائي راقي، رائحته شرقية زهرية جميلة وساحرة، عبيره فواح وقوي الثبات ويضيف لمسات أنيقة ملفتة للأنظار ولا تقاوم وتدوم لفترات طويلة. أطلقته شانيل عام ٢٠٠٥ ويعتبر من أهم الإصدارات ويمتاز برائحة ساحرة مفعمة بالحيوية وتعطي إحساس بالنعومة والدفء ويأتي عطر شانيل الور سنشوال في زجاجة شفافة مستطيلة الشكل ١٠٠ مل وبغطاء معدن نبيتي اللون فيضفي جمالا وسحرا على لون العطر، وتركيزه أو دي برفيوم، ويتكون من: مقدمة العطر: مقدمة العطر : البرغموت والباتشولي واليوسفي والفلفل الوردي قلب العطر: قلب العطر : زهور السوسن والياسمين والورد وفواكه مجففة قاعدة العطر: قاعدة العطر : التوابل والأخشاب ونجيل الهند والفانيليا والعنبر يتميز عطر شانيل الور سنشوال برائحة ساحرة ورقيقة تضفي جمالا للمرأة ويعكس أناقتها وأنوثتها فيفتتح العطر رائحة الحمضيات المنعشة و ينقلنا بسلاسة ونعومة لقلب من الزهور والفواكه ويختم بالأخشاب والعنبر والفانيليا فترسم هالة من الروائح الخيالية الساحرة ويتم استخدامه في جميع الأوقات والمناسباتآ	Buy Chanel Allure Sensual for Women Eau de Parfum - 100 ml online - Easy Click Chanel Allure Sensual is a sophisticated women's perfume. It has a beautiful and charming oriental floral scent. Its fragrance is fragrant and has a strong consistency. It adds elegant touches that catch the eye and are irresistible and last for long periods. See description for more information Chanel Allure Sensual is a sophisticated women's perfume. It has a beautiful and charming oriental floral scent. Its fragrance is fragrant and has a strong consistency. It adds elegant touches that catch the eye and are irresistible and last for long periods. It was launched by Chanel in 2005 and is considered one of the most important releases. It is characterized by a charming, lively scent that gives a feeling of softness and warmth. Chanel Allure Sensual perfume comes in a 100 ml transparent rectangular bottle with a burgundy metal cap, adding beauty and charm to the color of the perfume, and its concentration as eau de parfum. It consists of: Top notes: Top notes: bergamot, patchouli, mandarin, and pink pepper Heart of perfume: Heart of perfume: iris, jasmine, rose and dried fruits Base notes: Base notes: spices, woods, vetiver, vanilla and amber Chanel Allure Sensual perfume is characterized by a charming and delicate scent that adds beauty to a woman and reflects her elegance and femininity. The fragrance opens with a refreshing citrus scent and transports us smoothly and softly to a heart of flowers and fruits. It concludes with woods, amber and vanilla, creating an aura of enchanting imaginative scents. It is used at all times and occasions.

Table 7: Case study in Arabic. Included are both the original German texts and their English translations.