LLMs for Easy Language Translation: A Case Study on German Public Authorities Web Pages

Thorben Schomacker[1,2][0009−0002−3577−341X], Burak Tinman[1],
Chris Biemann[2][0000−0002−8449−9624], and
Marina Tropmann-Frick[1][0000−0003−1623−5309]

[1] Hamburg University of Applied Sciences, 20099 Hamburg, Germany
[2] University of Hamburg, 22527 Hamburg, Germany
thorben.schomacker@haw-hamburg.de

Abstract. This paper examines the use of Large Language Models (LLMs) for the intralingual translation of documents from standard German to German Easy Language (*Leichte Sprache*). We use open-weight models, from the Llama 3 family, with less than ten billion parameters. Additionally, we employ parameter-efficient fine-tuning (QLoRA) to adapt the LLMs to the requirements of Easy Language. For this purpose, we introduce a new data set (*ELGEPA*)[3], which is a parallel corpus of governmental documents in standard German and Easy Language with additional metadata, obtained from all German federal states and their capitals In our experiments, a fine-tuned Llama 3.1-8B-Instruct model achieved a SARI score of 41 and Flesch Reading Ease of 69. Outperforming GPT-4o and indicating that this type of model can deliver promising text quality in Easy Language.

Keywords: LLM · text simplification · Easy Language · Leichte Sprache · Governmental texts · intralingual translation

1 Introduction

The increasing importance of accessible digital communication calls for comprehensible language, e.g. Easy Language. Leichte Sprache (German Easy Language) is a simplified variant of German, characterized by simple sentence structures, clear wording and a comprehensibility-enhanced text structure. There are various international equivalents to Leichte Sprache, e.g., Easy Read (UK) or Easy-to-Read (USA, Australia) [9]. German public authorities are obliged to

[3] Source code to obtain the dataset: github.com/minds-hh/ger-gov-easy-lang

communicate in a comprehensible way [13]. Many of Germany's federal states already use translation services that convert complex articles into Easy Language on their websites. Even though some already use automated solutions, most translation services are conducting manual translations [3].

Large Language Models (LLMs) can provide a remedy by replacing this manual effort with machine translations. Open-weight models with a manageable number of parameters – i.e., models with fewer than ten billion parameters – offer decisive advantages over closed-weight alternatives such as GPT-4 or Claude Sonnet. This paper deals with the application of parameter-efficient fine-tuning (PEFT) methods, such as Low-Rank Adaptation (LoRA) or quantized LoRA (QLoRA), an efficient method of fine-tuning, to selected Llama 3 models with less than ten billion parameters in order to specialize them specifically for the generation of Easy Language texts. By adapting the pre-trained Llama 3 models using LoRA to a self-compiled data set of existing texts by public authorities in Easy Language, the ability of the model to generate coherent, grammatically correct texts that comply with the rules of Easy Language is to be tested.

2 Related Work

Easy Language translation can be considered as *automatic text simplification*, where a lexical and structural complex text is translated into a simplified version, while preserving its meaning. German text simplification research has recently focused on machine learning-based methods (see [31,23] for a recent overview). Translation into Easy Language poses a particular challenge, because it is more than just a simplified version of a language; it follows specific rules and conventions. It is therefore important that models not only correctly implement the semantic content, but also these formal aspects of Easy Language. There are various sets of rules [8,7,28]. With the DIN SPEC 33429 [6], finalized in March 2025, likely becoming the ruling guideline in the future. The main problem with all machine learning methods is the data basis, which is scarce in the field of German for parallel corpora between complex texts and Easy Language. Some research work has dealt with the creation of a general parallel corpus in German [20,12,31,34], but data sets for governmental texts are few to none. In order to deal with the scarcity of data, however, there are approaches to train the language models with (semi)-synthetic data [21] or to pursue a style-specific pre-training approach [2].

Finally, the evaluation of automated text simplification systems is an under-researched field that is complicated by the subjectivity of the task and lack of standards, which makes the development of suitable metrics and reference data, especially for non-English languages, necessary [14]. Automated metrics such as BLEU [29], ROUGE [22] or SARI [37] are often used, but they cannot fully capture the special requirements of Easy Language. Operationalizing part of the parts of Easy Language in a form evaluation has been done, but is still in its infancy [16,24,32,30,15].

3 Methods

Since parallel corpora of governmental texts in German, including Easy Language, are hardly available[4], the first step was to gather relevant data. In the next step, appropriate models were identified that were suitable for fine-tuning. During the fine-tuning itself, the selected models were trained using the PEFT approach and the results were then validated and checked using various metrics.

3.1 Dataset Creation

The first step was to check the 16 websites of the federal states of Germany. As only 3 of these offered the possibility of extracting a parallel corpus, the search was extended to the websites of the federale state capitals (excluding the city states: Berlin, Bremen and Hamburg). With the exception of Magdeburg, which uses a live AI-controlled translation into Easy Language and is therefore difficult to extract, the remaining possible state capitals are the basis for our dataset: Easy Language documents of GErman Public Authorities (*ELGEPA*) shown in Table 1. In summary, 609 documents with standard and Easy Language were extracted from the 8 root websites using Scrapy[5]. We further expanded the data set with the Cologne subset from [36]. The subject areas of the texts focus primarily on help, tips, news or regional content.

Furthermore, if given in the source, we added metadata to each sample: Whether the text was generated in Easy Language by an AI, the corresponding author of the text, the proofreader of the text, which form of address[6] was used in the text and the subject area covered by the text. Subsequently, the dataset was divided into a training and test split in a ratio of 87.5% to 12.5% so that the training dataset contains at least 500 training copies and the test dataset has enough copies to be able to check the performance of the models in a meaningful way. The data set also had to be converted into a machine-readable format for subsequent fine-tuning. We used the ShareGPT[7] format, which is particularly suitable for multi-turn conver-

Source	**# docs**
hannover.de	272
hamburg.de	103
stadt-koeln.de	82
dresden.de	67
stuttgart.de	44
stadt.muenchen.de	28
hessen.de	5
saarbruecken.de	4
ms.niedersachsen.de	4
total	609

Table 1. Number of documents per source in our dataset

[4] There is only a subset from for the city of Cologne [36]

[5] `https://www.scrapy.org`, last access 16.04.2025

[6] Gemini 1.5 Flash, was prompted to choose "Duzen", "Siezen" or a mix of both to determine the text's form of address. This feature of our data was discussed in [33]

[7] It represents conversations as JSON arrays of message objects. Each message includes a 'from' and a 'value'-field, e.g. `{'conversations': [{'from': 'human', 'value': <question>}, {'from': 'gpt','value': <response>}]}`. For more information and examples see `https://huggingface.co/datasets/anon8231489123/ShareGPT_Vicuna_unfiltered`, last access 16.04.2025

sations, but can also be used for first-turn conversations to support a prompt-response structure.

3.2 Model Selection

Once the data set had been transformed into the ShareGPT format, the fine-tuning could begin. Since the data basis is scarce, especially in the field of Easy Language, fine-tuning or more specialized methods such as PEFT, e.g. LoRA or QLoRA [1], can also be used to train a powerful model with less data in a resource-saving manner. Other selection criteria included that the models were decoder-only and already instruct-based. A style training approach similar to that of Anschütz et al [2] was pursued, without ultimately embedding the decoder model in a sequence-to-sequence model.

Meta's Llama 3 model family was selected because it has received widespread recognition in the research community and is considered to be highly performant. This is particularly evident on Huggingface, where the aforementioned Llama models have the most downloads in the text generation area (as of 08.01.2025). Within this family, three models were selected for fine-tuning: Llama 3.1 8B-Instruct [26], Llama 3.2 3B-Instruct, and Llama 3.2 1B-Instruct [27]. Models 3B and 1B are the first Meta models with a size of less than five billion parameters. This range of parameters allows us to investigate the influence of model size on the quality of the generated Easy Language texts.

3.3 Fine-Tuning

To fine-tune the models, Unsloth.ai[8] [10] was employed, which uses Triton [35] to implement backpropagation in LoRA training processes. Using Triton makes it possible to minimize the number of floating-point operations (FLOPs) during gradient descent, which has a positive impact on training speed [39].

Using the LoRA [17] method enables more efficient fine-tuning by freezing the pre-trained weights and training only a few additional parameters. This reduces the number of trainable parameters and GPU memory requirements during training, while maintaining or even improving the model quality in comparison to full fine-tuning. QLoRA [11] is a development of LoRA that enables the fine-tuning of extremely large language models on GPUs with limited memory by quantizing the pre-trained model to 4 bits and further memory optimizations. In contrast to LoRA, QLoRA not only reduces the number of trainable parameters but also the memory requirements of the base model itself, enabling even greater fine-tuning efficiency.

Based on this, the appropriate settings were made for the selected Llama 3 models. We configured the 4-bit quantized Llama 3.1-8B-Instruct model, to have

[8] We adapted pre-built Notebooks (colab.research.google.com/notebook) from Unsloth (unsloth.ai) for our fine-tuning. Additionally, Unsloth offers pre-built 4-bit quantized Llama 3 models, leveraging bitsandbytes, which enables k-bit quantization[25]. This also allows for fine-tuning using the QLoRA method [5].

a maximum sequence length of 8192 tokens for training, which allowed most of the dataset elements to fit into the training context without requiring too much resource-intensive training. The next step involved configuring the QLoRA. All of this was adopted from the template, as the default settings were already sufficient for this use case.

The data was then prepared for training, by transferring it to the Llama 3 model's chat template and enriching it with as system prompt. This served as an instruction for creating the texts in Easy Language and was embedded in all created prompts. The system prompt consists of a general instruction, a set of rules for translating into Easy Language, and a one-shot prompt example. The set of rules is taken from Regelwerk[9], which refers to the set of rules from [28]. The training[10] consisted of a total of 4 training epochs with a total batch size of 64 and 32 steps. However, the number of 511 training samples is still sufficient for training QLoRAs.

	Reference	**Base 3.2-1B**	**Base 3.2-3B**	**Base 3.1-8B**	**QLoRA 3.2-1B**	**QLoRA 3.2-3B**	**QLoRA 3.1-8B**	GPT 4o
FR Ease	$\uparrow$ 71.105	51.04	58.84	62.745	61.19	56.89	**69.38**	43. 91
FK Grade	$\downarrow$ 5.65	10.05	7.95	6.8	7.75	8.2	**6.15**	11.3
WSF	$\downarrow$ 3.6	6.75	5.45	4.8	5.7	6.0	**4.2**	7.65
SARI	$\uparrow$ -	35.47	38.86	38.37	39.79	40.16	**41.31**	39.58
BERTScore	$\uparrow$ -	0.63	0.65	0.65	0.64	0.64	0.66	**0.68**

Table 2. Median values of readability indexes: Flesch Reading Ease (FR Ease) $\uparrow$, Flesch-Kincaid Grade (FK Grade) $\downarrow$, Wiener Sachtextformel (WSF) $\downarrow$ and reference-based metrics: SARI $\uparrow$, BERTScore $\uparrow$. For metrics with an $\uparrow$, higher values are better results. Vice versa, $\downarrow$ indicates that lower values are better.

3.4 Evaluation

For the evaluation of machine-translated texts, there are many different metrics that check the aspects of text quality for texts from automated text simplification systems and can therefore also indirectly evaluate texts in Easy Language.

The Flesch Reading Ease Index [19], quantifies the readability of a text on a scale of 0-100. Higher values indicate greater ease of reading, based on the average sentence length and the number of syllables per word. Similarly, the Flesch-Kincaid Grade Index [18] indicates the approximate school level required to understand the text on a scale from 0-18. Lower values indicate easier comprehension. Wiener Sachtextformel (Variant 4) [4], is used to assess the comprehensibility of factual texts on a scale of 4-15. Among other things, it takes into account the average sentence length and the number of words with more than three syllables. The lower the value, the easier the text is to understand. The SARI (System output Against References and Input) metric [37] evaluates the quality of a simplified text on a scale of 0-100 by analyzing the degree of

[9] de.wikipedia.org/w/index.php?title=Leichte_Sprache, last access 16.04.2025
[10] Due to an error, the Munich subset is missing from the training and test data

addition, deletion and retention of information compared to the original text. It therefore measures how well a simplification system retains the meaning and essential content. The higher the value, the better the performance. BERTScore [38] uses contextual embeddings[11] to assess the semantic similarity between the generated and the reference text on a scale of 0-1. This enables a more differentiated evaluation of text quality compared to other metrics. The higher the score, the better the performance.

To evaluate the generated texts, a test data set with 74 samples from our data set was used as a reference. The standard texts were used as input together with the system prompt for the various Llama 3 models. The results in Table 2 show that the reference texts have better readability scores, which is reflected in a higher Flesch Reading Ease and lower Flesch-Kincaid Grade as well as in the Wiener Sachtextformel values. This indicates that the reference texts are easier to understand. In particular, the QLoRA 3.1-8B model best approximates the readability values of the original text on average (see also Table 2). In summary, the results nevertheless indicate that fine-tuning using QLoRA can be an effective method for optimizing LLMs for the generation of texts in Easy Language. In particular, the QLoRA model Llama-3.1-8B-Instruct proves to be more efficient in terms of readability and retention of semantic content as compared to the other models. Furthermore, compared to GPT-4o, prompted on the same task, our model considerably outperformed in terms of readability and SARI value. While achieving comparable results for the BERTScore. Particularly, the gap in the readability scores in comparison to the reference texts from our dataset show potential for improvement. This could be the result of a tendency to retain long words from the standard language text into the Leichte Sprache one, which we observed in some of the generated texts.

4 Conclusions and Outlook

Our results show that fine-tuning LLMs with QLoRA can be a promising approach for automated translation into Easy Language. In particular, it was shown that the QLoRA-fine-tuned Llama 3.1-8B-Instruct model achieved the best results as compared to the base models, both in terms of readability and retention of semantic content. The larger model showed an improved adaptation to the specific requirements of Easy Language. It should also be noted that all trained models did not quite reach the readability values of the reference texts, but were able to perform a certain degree of text simplification according to the system prompt. This can also be seen in the higher reference metrics in the QLoRA models as compared to the basic models.

Nevertheless, it became evident that the smaller models with less than 8 billion parameters are not yet fully capable of consistently generating meaningful text in Easy Language. In particular, more complex system prompts or instructions cannot always be adhered to. Furthermore, better and more powerful open

[11] In our case from `bert-base-multilingual-cased`

weight or source models are expected in the future, which may be more consistent in terms of text generation and incorporating additional Easy Language data could also lead to an increase of the performance. In addition, clear regularizations for Easy Language would be desirable, as well as specialized ways to measure the quality of Easy Language texts. However, as mentioned in [6], these are already in progress. In addition, prompt engineering, based on the existing rules on Easy Language, could lead to a performance increase of the models we discussed and similar ones.

References

1. Anisuzzaman, D.M., Malins, J.G., Friedman, P.A., Attia, Z.I.: Fine-Tuning Large Language Models for Specialized Use Cases. Mayo Clinic Proceedings: Digital Health **3**(1), 100184 (Mar 2025). https://doi.org/10.1016/j.mcpdig.2024.11.005
2. Anschütz, M., Oehms, J., Wimmer, T., Jezierski, B., Groh, G.: Language models for German text simplification: Overcoming parallel data scarcity through style-specific pre-training. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Findings of the Association for Computational Linguistics: ACL 2023. pp. 1147–1158. Association for Computational Linguistics, Toronto, Canada (Jul 2023). https://doi.org/10.18653/v1/2023.findings-acl.74, https://aclanthology.org/2023.findings-acl.74/
3. Asghari, H., Hewett, F., Züger, T.: On the Prevalence of Leichte Sprache on the German Web. In: Proceedings of the 15th ACM Web Science Conference 2023. pp. 147–152. WebSci '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3578503.3583599
4. Bamberger, R., Vanecek, E.: Lesen - Verstehen - Lernen - Schreiben. Die Schwierigkeitsstufen von Texten in Deutscher Sprache. Jugend u. Volk Sauerlaender, Vienna, Austria (1984)
5. Belkada, Y., Dettmers, T., Pagnoni, A., Gugger, S., Mangrulkar, S.: Making LLMs even more accessible with bitsandbytes, 4-bit quantization and QLoRA. https://huggingface.co/blog/4bit-transformers-bitsandbytes (Mar 2025)
6. Beyer, I., Nietzio, A., Sieghart, S., Borinski, U., Krull, R., Bode-Welch, B., Schindofski, R., Pasquavaglio, S., Hütter-Songallo, M., Bilk, J., Baltscheit, K., Pool, A.J., Schiffler, I., Szymanowicz, T., Wagner, Y., Koehler, S., Meyer, D., Wenzl, J., Gutermuth, S., Herdejürgen, K., Seidel, A., Wallner, S., Klanke, M., Zornow, S., Treptow, U., Schrübbers, C., Weiland, C., Lotze, T., Nölting, M., Wegge, K.P., Handrick, A., Müller, S., Brandes, J., Lasch, A., Gorbach, R.P., Wahl, M., Lindmeier, B., Heidelberger, J., Nolte, H.: DIN SPEC 33429 - Empfehlungen für Deutsche Leichte Sprache (Mar 2025)
7. Bredel, U., Maaß, C.: Leichte Sprache Theoretische Grundlagen, Orientierung Für Die Praxis. Sprache Im Blick, Dudenverlag, Berlin, Germany (2016)
8. Bundesministerium für Arbeit und Soziales: Leichte Sprache Ein Ratgeber (Apr 2014)
9. Chinn, D., Buell, S.: Easy Language in the UK. In: Lindholm, C., Vanhatalo, U., Hansen-Schirra, S., Maaß, C. (eds.) Handbook of Easy Languages in Europe, pp. 623–657. No. 8 in Easy–Plain–Accessible, Frank & Timme, Berlin, Germany, 1 edn. (2021). https://doi.org/10.26530/20.500.12657/52628
10. Daniel Han, M.H., team, U.: Unsloth (2023)
11. Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L.: QLORA: Efficient finetuning of quantized LLMs. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. pp. 10088–10115. NIPS '23, Curran Associates Inc., Red Hook, NY, USA (Dec 2023)
12. Ebling, S., Battisti, A., Kostrzewa, M., Pfütze, D., Rios, A., Säuberli, A., Spring, N.: Automatic Text Simplification for German. Frontiers in Communication **7**, 706718 (Feb 2022). https://doi.org/10.3389/fcomm.2022.706718
13. Gille, M., Schomacker, T., von der Hülls, J., Tropmann-Frick, M.: Der Einsatz von Neural Language Models für eine barrierefreie Verwaltungskommunikation: Anforderungen an die automatisierte Vereinfachung rechtlicher Informationstexte.

In: Proceedings of 6. Fachtagung Rechts- und Verwaltungsinformatik (RVI 2023). pp. 144–158. Gesellschaft für Informatik e.V., Bonn (26-27 October 2023). https://doi.org/10.18420/rvi2023-013
14. Grabar, N., Saggion, H.: Evaluation of automatic text simplification: Where are we now, where should we go from here. In: Estève, Y., Jiménez, T., Parcollet, T., Zanon Boito, M. (eds.) Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale. pp. 453–463. ATALA, Avignon, France (6 2022), https://aclanthology.org/2022.jeptalnrecital-taln.47/
15. Hewett, F., Asghari, H., Stede, M.: Elaborative Simplification for German-Language Texts. In: Kawahara, T., Demberg, V., Ultes, S., Inoue, K., Mehri, S., Howcroft, D., Komatani, K. (eds.) Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue. pp. 29–39. Association for Computational Linguistics, Kyoto, Japan (Sep 2024). https://doi.org/10.18653/v1/2024.sigdial-1.3
16. Hewett, F., Stede, M.: Automatically evaluating the conceptual complexity of German texts. In: Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021). pp. 228–234. KONVENS 2021 Organizers, Düsseldorf, Germany (2021)
17. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models. In: Proceedings of the International Conference on Learning Representations. Online (Jan 2022)
18. Kincaid, J.P., Fishburne, Jr., Robert P., R., Richard L., C., Brad S.: Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel:. Reports - Research ED108134, Defense Technical Information Center (Feb 1975). https://doi.org/10.21236/ADA006655
19. Kincaid, J.P., Aagard, J.A., O'Hara, J.W., Cottrell, L.K.: Computer readability editing system. IEEE Transactions on Professional Communication **PC-24**(1), 38–42 (Mar 1981). https://doi.org/10.1109/TPC.1981.6447821
20. Klaper, D., Ebling, S., Volk, M.: Building a German/Simple German Parallel Corpus for Automatic Text Simplification. In: Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR 2013), Sofia, Bulgaria, 8 August 2013. pp. 11–19. Sofia, Bulgaria (Aug 2013). https://doi.org/10.5167/uzh-78610
21. Klöser, L., Beele, M., Schagen, J.N., Kraft, B.: German Text Simplification: Finetuning Large Language Models with Semi-Synthetic Data. In: Chakravarthi, B.R., B, B., Buitelaar, P., Durairaj, T., Kovács, G., García Cumbreras, M.Á. (eds.) Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion. pp. 63–72. Association for Computational Linguistics, St. Julian's, Malta (Mar 2024)
22. Lin, C.Y.: ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out. pp. 74–81. Association for Computational Linguistics, Barcelona, Spain (Jul 2004), https://aclanthology.org/W04-1013/
23. Madina, M., Gonzalez-Dios, I., Siegel, M.: Easy-to-Read in Germany: A Survey on its Current State and Available Resources. In: Proceedings of the 10th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics. Poznań, Poland (Apr 2023)
24. Madina, M., Gonzalez-Dios, I., Siegel, M.: Towards Reliable E2R Texts: A Proposal for Standardized Evaluation Practices. In: Miesenberger, K., Peňáz, P., Kobayashi,

M. (eds.) Computers Helping People with Special Needs. pp. 224–231. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-62849-8_28
25. Matthew Douglas: Bitsandbytes 0.45.5. bitsandbytes foundation (Apr 2025)
26. Meta AI: Introducing Llama 3.1: Our most capable models to date. https://ai.meta.com/blog/meta-llama-3-1/ (Jul 2024)
27. Meta AI: Llama 3.2: Revolutionizing edge AI and vision with open, customizable models. https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/ (Sep 2024)
28. Netzwerk Leichte Sprache: Die Regeln für Leichte Sprache (2022)
29. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Isabelle, P., Charniak, E., Lin, D. (eds.) Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. pp. 311–318. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA (Jul 2002). https://doi.org/10.3115/1073083.1073135, https://aclanthology.org/P02-1040/
30. Schomacker, T., Anschütz, M., Stodden, R., Groh, G., Tropmann-Frick, M.: Overview of the GermEval 2024 shared task on statement segmentation in German easy language (StaGE). In: Schomacker, T., Anschütz, M., Stodden, R. (eds.) Proceedings of GermEval 2024 Shared Task on Statement Segmentation in German Easy Language (StaGE). pp. 1–14. Association for Computational Lingustics, Vienna, Austria (Sep 2024), https://aclanthology.org/2024.germeval-1.1/
31. Schomacker, T., Gille, M., Tropmann-Frick, M., von der Hülls, J.: Data and Approaches for German Text simplification – towards an Accessibility-enhanced Communication. In: Georges, M., Herygers, A., Friedrich, A., Roth, B. (eds.) Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023). pp. 63–68. Association for Computational Lingustics, Ingolstadt, Germany (Sep 2023)
32. Schomacker, T., Gille, M., Tropmann-Frick, M., von der Hülls, J.: Self-regulated and Participatory Automatic Text Simplification. In: Jabbar, M.A., Tiwari, S., Ortiz-Rodríguez, F., Groppe, S., Bano Rehman, T. (eds.) Applied Machine Learning and Data Analytics. pp. 264–273. Communications in Computer and Information Science, Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-55486-5_19
33. Schomacker, T., Tropmann-Frick, M., Gille, M.: "Duzen oder Siezen?" – Forms of address in Easy Language webpages of German public authorities (Mar 2025)
34. Stodden, R., Momen, O., Kallmeyer, L.: DEplain: A German Parallel Corpus with Intralingual Translations into Plain Language for Sentence and Document Simplification. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 16441–16463. Association for Computational Linguistics, Toronto, Canada (Jul 2023). https://doi.org/10.18653/v1/2023.acl-long.908
35. Tillet, P., Kung, H.T., Cox, D.: Triton: An intermediate language and compiler for tiled neural network computations. In: Proceedings of the 3rd ACM SIGPLAN International Workshop on Machine Learning and Programming Languages. pp. 10–19. MAPL 2019, Association for Computing Machinery, New York, New York, USA (Jun 2019). https://doi.org/10.1145/3315508.3329973
36. Toborek, V., Busch, M., Boßert, M., Bauckhage, C., Welke, P.: A new aligned simple German corpus. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 11393–11412. Association for Computational Linguistics,

Toronto, Canada (Jul 2023). https://doi.org/10.18653/v1/2023.acl-long.638, https://aclanthology.org/2023.acl-long.638/

37. Xu, W., Napoles, C., Pavlick, E., Chen, Q., Callison-Burch, C.: Optimizing statistical machine translation for text simplification. Transactions of the Association for Computational Linguistics **4**, 401–415 (2016). https://doi.org/10.1162/tacl_a_00107, https://aclanthology.org/Q16-1029/
38. Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y.: BERTScore: Evaluating Text Generation with BERT. In: Proceedings of the Eighth International Conference on Learning Representations. Addis Ababa, Ethiopia (Apr 2020)
39. Zheng, Y., Zhang, R., Zhang, J., Ye, Y., Luo, Z.: LlamaFactory: Unified efficient fine-tuning of 100+ language models. In: Cao, Y., Feng, Y., Xiong, D. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). pp. 400–410. Association for Computational Linguistics, Bangkok, Thailand (Aug 2024). https://doi.org/10.18653/v1/2024.acl-demos.38, https://aclanthology.org/2024.acl-demos.38/

A System-Prompt for Fine-Tuning and Evaluation

Du bist ein Sprachassistent, der normalen deutschen Text in Leichte Sprache übersetzt. Halte dich strikt an die folgenden Regeln:

1. Verwende kurze Sätze. Jeder Satz enthält nur eine Aussage. 2. Schreibe nur Aktivsätze im Format Subjekt-Verb-Objekt. 3. Vermeide den Konjunktiv. Schreibe nur in der Wirklichkeitsform. 4. Ersetze den Genitiv durch präpositionale Fügungen mit „von“. 5. Nutze keine Synonyme oder Sonderzeichen. 6. Formuliere Verneinungen positiv, wenn möglich. 7. Verwende präzise Mengenangaben nur in Ausnahmefällen; ersetze sie durch Begriffe wie „viel“ oder „wenig“. Jahreszahlen werden durch Ausdrücke wie „vor langer Zeit“ ersetzt. 8. Leichte Sprache ist keine Kindersprache. Verwende „Du“ und „Sie“ korrekt wie in der Standardsprache.

Rechtschreibregeln: 1. Verwende Bindestriche oder Mediopunkte, um Zusammensetzungen zu verdeutlichen (z. B. Welt-All oder Welt·all). 2. Nutze keine durchgehenden Großbuchstaben oder Kursivschrift. 3. Schreibe jeden Satz in eine eigene Zeile.

Regeln für den Inhalt: 1. Vermeide abstrakte Begriffe oder erkläre sie mit anschaulichen Beispielen. 2. Vermeide bildhafte Sprache. 3. Erkläre Fremd- und Fachwörter bei der ersten Erwähnung. 4. Schreibe Abkürzungen beim ersten Vorkommen aus.

Gestaltungsregeln: 1. Gestalte den Text übersichtlich. Jeder Satz steht in einer eigenen Zeile und linksbündig im Flattersatz. 2. Verwende Aufzählungspunkte für Listen. 3. Bilder und Text sollen getrennt bleiben. Bilder können zur Veranschaulichung hinzugefügt werden.

Beispiel: Normaler Text: 'Bringen Sie für Ihre stationäre Aufnahme in eine unserer Kliniken bitte die notwendigen Formulare, sewie den Einweisungsschein Ihres Arztes, die Krankenversicherungskarte und gültige Personalpapiere mit.'

Leichte Sprache: 'Sie sind krank?

Sie müssen ins Kranken-Haus zu einer Untersuchung? Oder zu einer Operation?

Dann bringen Sie bitte diese Papiere mit ins Kranken-Haus:

Den Zettel von Ihrem Haus-Arzt. Auf diesem Zettel steht, dass Sie ins Kranken-Haus müssen. Die Karte der Kranken-Versicherung. Ihren Ausweis.

Bitte schauen Sie vorher in Ihrem Ausweis ein Datum nach.

Unter dem Punkt „Ablaufdatum“ steht ein Datum mit einer Jahreszahl.

Ist das dort angegebene Datum schon vorbei?

Dann gilt dieser Ausweis nicht mehr.

Sie müssen dann einen anderen Ausweis mitbringen,

z. B. den Reisepass.'