Developing A German Document-Level Parallel Dataset For Automatic Text Simplification To Generate Easy Language

Vivien Jiranek **Stefan Hillmann**
Quality and Usability Lab, Technische Universität Berlin
Straße des 17. Juni 135, 10623 Berlin, Germany
v.jiranek@protonmail.com stefan.hillmann@tu-berlin.de

Abstract

A lack of suitable datasets limits research on automatically generating German Easy-to-Read Language (GE2R), known as Leichte Sprache in Germany, from regular German. To address this, we introduce the EasyGerman dataset, the largest collection of parallel documents in regular German and professionally translated German Easy Language in the news domain. We demonstrate the feasibility of generating GE2R by fine-tuning pre-trained mt5 and mBART Large Language Models on this dataset using transfer learning. These models are evaluated using various metrics to establish a reliable benchmark for future research. Additionally, a human evaluation involving members of the target group was conducted to assess the quality and readability of the highest-scoring generated texts. Metric scores show that all models effectively generate simplified German in line with GE2R guidelines. The best approach produced texts rated as easy to read and consistent with GE2R principles, although there were mixed results in content representation and identified issues with logical consistency. Our findings highlight the value of the EasyGerman dataset for advancing research on automatic GE2R generation using neural models.

1 Introduction

Easy-to-Read language, also known as Easy Language, is an internationally recognized simplified language variation. It has a low linguistic complexity that is designed to accommodate people with low literacy, promoting inclusive communication. It distinguishes itself from other simplified language variations such as Plain Language (PL), since it was deliberately tailored to account for the communicative needs of disabled people. Therefore, the use of GE2R is mandated by German Law (Bundesministerium der Justiz, 2002) for information published by public authorities, thus ensuring the supply of accessibility-enhanced information to the general public. Currently, the high demand for translations from regular German (RG) to GE2R exceeds the capacity of available professional translators. Yet research on the generation of automatic German Text Simplification is scarce due to a lack of suitable parallel data, especially involving neural approaches. Thus, we developed the EasyGerman dataset, the largest document-level parallel dataset featuring news data in RG and GE2R to this date. It incorporates professionally translated data that was regularly published by the German public broadcaster Mitteldeutscher Rundfunk since 2021. Furthermore, we developed four approaches to benchmark the dataset and determine the feasibility of automatically generating GE2R. We use large language models leveraging transfer learning to overcome the lack of pre-trained models and available document-level parallel data for pre-training for this German Text Simplification task. Their performance was evaluated using a variety of automatic metrics carefully selected for this purpose. Finally, we conducted a quality assessment of the output texts of the best-performing approach using an online survey. These three contributions and the related results are presented in this paper.

2 Related Work

The use and prevalence of languages on a global scale directly relates to the amount of publicly available data to build additional datasets for research. Prominent sources such as Simple Wikipedia, which offers simplified Wikipedia articles on various topics, are unavailable in German. Therefore, gathering data to build German datasets for more specialized NLP tasks such as text simplification on a similar scale as the the D-Wikipedia dataset (Sun et al., 2021), is currently unattainable (Rios et al., 2021). In recent years, several datasets have been built incorporating different language variations and simplification levels. Though most of them feature GE2R, emphasizing its significance in Ger-

Proceedings of the 21st Conference on Natural Language Processing (KONVENS), Volume 2: Workshops, pages 142–168, 2025

man text simplification research, many do not exclusively feature or were intended for research on GE2R.

In 2023, a survey by Madina et al. summarized the current state of research on German Text Simplification, with a particular focus on GE2R. We have extended their survey by several more recent datasets, such as the EasyGerman dataset, presented in Table 1. This table provides a concise overview of key dataset properties, organized by their release date. In this context, "Simplified German" refers to tailored simplified language variations, "General Simplicity" refers to RG at lower complexity levels, and "German Plain Language" is abbreviated as GPL. The datasets in Table 1 reveal key trends and challenges in German Text Simplification, notably a shift from sentence-level to document-level alignments. The comparison also highlights difficulties in developing parallel datasets, with significant variation in dataset sizes, alignment granularity, and featured languages. While some datasets provide extensive resources, others are smaller or lack quantitative details, highlighting the uneven availability of resources and the need for more standardized datasets. Notably, the EasyGerman dataset stands out based on its provision of parallel data, exclusive document-level granularity, and its consistent focus on GE2R texts as the target language, as well as its comparatively large amount of aligned documents.

3 EasyGerman Dataset

The EasyGerman dataset[1] was developed as a document-level parallel dataset to enable further research on automatic German Text Simplification with a primary focus on the generation of GE2R.

3.1 Data Collection

The dataset comprises a total of 3,832 news articles, originally published by the German public broadcaster Mitteldeutscher Rundfunk (MDR) between July 2021 and April 2024. Each professionally translated article in GE2R is linked to a source article in RG by the publisher. This predetermined alignment is maintained throughout the dataset. The full dataset provides a total of 1,916 sample-pairs. The news articles cover various topics including politics, economics, sports, weather, and cultural issues. Articles were primarily sourced through regular web scraping of the MDR's **W**eekly **N**ews website in GE2R (MDR-WN)[2]. Additionally 538 articles were obtained from MDR's **N**ews **A**rchive (MDR-NA)[3].

To further expand the dataset, 1,906 articles published between June 2021 and February 2023 were retrieved from the **I**nternet **A**rchive (IA)[4]. Articles predating June 2021 were excluded due to the absence of links between GE2R texts and their RG counterparts. Table 2 provides the number of sample-pairs available per source and year.

To avoid the possibility of copyright claims, the text files containing the gathered URLs of the data pairs as well as the implemented Python scripts for web scraping and building the dataset are made available to build the dataset on demand. Since the published articles are publicly available when given the proper URL, but the URLs of published articles are removed at the beginning of every new week on the MDR weekly news website and after three months predating the current month on the MDR News archive, these text files are essential to access and incorporate previously released data pairs.

Both the raw HTML content and relevant metadata from the HTML header are automatically collected from each news article's webpage during the web-scraping process. The metadata provides insights on the domain and type of information gathered and its selection was inspired by the metadata collection of Battisti et al. 2020. It encompasses the title, description, keywords in GE2R, URL, and date of publishing as well as the given document identifier, linking the content of each article with a downloaded file, the cleaned text written to said file, and the raw HTML-content of each scraped webpage.

3.2 Alignment

A document-level alignment was chosen over the more common sentence-level alignment due to the characteristics of the news domain and the nature of GE2R as a language subset. Document-level simplifications address the complexity of such real-world applications by preserving the discourse structure of both input and output documents (Blinova et al., 2023). This allows neural models to

[1] https://github.com/vjiranek/EasyGermanDataset

[2] https://www.mdr.de/nachrichten-leicht/nachrichten-in-leichter-sprache-114.html

[3] https://www.mdr.de/nachrichten-leicht/rueckblick/index.html

[4] https://web.archive.org/web/20240000000000*/https://www.mdr.de/nachrichten-leicht/

Table 1: German Text Simplification Datasets

Dataset	Alignment	Featured Languages	Quantity
Klaper et al. (2013)	Sentence	GE2R	256 aligned documents
LeiSa (2018)	None	GPL, GE2R and simplified German	639,826 words in GE2R, 779,278 words in GPL, and 350,872 words in simplified DE
Battisti and Ebling (2019)	Partially Document	GE2R	378 aligned documents and 5,461 monolingual documents
TextComplexityDE (2019)	Sentence	Simplified German	1000 sentences in total, 250 aligned sentences
LeiKo (2020)	None	GE2R	50,000 words
KED (2023)	Sentence	GE2R and GPL	3,698,372 words
APA (2020)	Sentence	CEFR levels A2, and B1	3,616 aligned sentences
Capito (2023)	Sentence	CEFR levels A1, A2, and B1	NA
20m (2021)	Document	General Simplicity	18,305 aligned summaries
GEASY (2021)	Sentence, Document	GE2R	289 documents
Toborek et al. (2022)	Sentence, Document	GE2R and GPL	708 aligned documents, 5,942 aligned sentences
Klexicon (2022)	Document	General Simplicity	2,898 aligned documents
SNIML (2022)	None	Simplified Languages including GE2R	147 German articles with 25,964 sentences
DEPLAIN (2023)	Sentence, Document	GPL	483 aligned documents, 147 aligned sentences
DE-Lite (2024)	Various	Simplified German and General Simplicity	NA
EasyGerman	Document	GE2R	1916 aligned documents

learn interdependencies within the text, which was expected to improve output quality. In contrast, sentence-level mappings were found to negatively impact results due to reduced text length, omissions, and reordering of topics (Säuberli et al., 2020). Automatic sentence mappings were deemed infeasible, as most tools are designed for translation and thus demonstrate shortcomings with simplified, paraphrased language. Since manual mappings were not expected to yield positive results and were also not considered viable due to the labor required, document-level alignment became the preferred choice to preserve the publisher's intended parallel structure and maintain high-quality output.

Table 2: Number of Incorporated News Articles per Source Website Published During Each Year

	Year				
Source	**2021**	**2022**	**2023**	**2024**	**Total**
MDR-WN	0	0	887	505	**1,392**
MDR-NA	0	0	538	0	**538**
IA	465	1,181	260	0	**1,906**
Total	**465**	**1,181**	**1,685**	**505**	**3,832**

3.3 Language Analysis

Automatically computed statistics on the language characteristics and complexity give insights on the linguistic properties of the dataset and highlight differences between the articles in GE2R and RG. The total number of letters, words, sentences, and vocabulary size within the full dataset, along with the average and median values for a single document is shown in Table 3. Our analysis follows the dataset assessment method in (Battisti et al., 2020).

The EasyGerman dataset contains approximately 6.39 million letters, 1.2 million words, and 96,000 sentences. Notably, only the letters of tokenized words were considered in the linguistic analysis, excluding whitespaces and other characters, to provide information on the complexity of the underlying language. The full dataset has a total vocabulary size of 71,912 unique words, with an average of 162 words per document. A comparison of lin-

guistic features between RG and GE2R indicates substantial differences in both their length and language complexity as seen in Table 3.

The size ratios between the two language variations show that GE2R documents are much shorter in word and letter count. The word ratio between GE2R and RG is about 1:2, with RG containing nearly twice as many words. The letter ratio is around 2:5, with RG contributing a larger share. Despite these differences, the total number of sentences is almost identical for both variations, indicating that while sentences in GE2R are shorter, the sentence count remains unchanged. On average, sentences in RG have 17 words, while those in GE2R have only 9 (see Table 4). Interestingly, the average letters per word is similar for both, with GE2R averaging 5 letters and RG 6, suggesting that word length is comparable despite the shorter sentences in GE2R.

Further analysis of the vocabulary size underscores the differences in language diversity and complexity. The median (*italic*) and average (**bold**) counts of linguistic properties can be seen in Table 5. The vocabulary used in the RG documents is almost twice as large as of the documents written in GE2R (on average 223 vs 101). These differences reflect the purposefully simplified nature of GE2R. Finally, it is noticeable that the dataset exhibits a right-skewed distribution, as indicated by the difference between the median and average values. The median values are consistently smaller than the averages, suggesting that a small subset of longer, more complex documents skews the overall statistics. This variability could have implications for machine-learning models trained on this data and should be considered during further use.

Table 3: Total Number of Linguistic Units

		Words		
	Letters	Total	Unique	Sentences
Full Dataset	6,388,728	1,196,192	71,912	96,263
Leichte **S**prache	1,908,071	389,896	17,707	46,293
Regular**G**erman	4,480,657	806,296	63,452	49,970
Ratio **GE2R**:**RG**	2:5	1:2	1:4	9:10

Table 4: Average Word and Sentence Length

	Full Dataset	GE2R	RG
Letters per Word	6	5	6
Words per Sentence	13	9	17

Table 5: **Average** and *Median* Number of Linguistic Units per Document

		Words		
	Letters	Total	Unique	Sentences
Full Dataset	**1668** *1257.0*	**313** *242.0*	**162** *124.0*	**26** *23.0*
GE2R	**996** *924.0*	**204** *189.0*	**101** *98.0*	**25** *22.0*
RG	**2339** *2100.5*	**421** *375.5*	**223** *205.0*	**27** *23.0*

4 Benchmarking

To automatically generate texts in GE2R from texts in RG, three different large language models using different task presets were fine-tuned for up to 20 epochs and then evaluated for their performance using the EasyGerman dataset. Different hyperparameter configurations were tested involving a range of different learning rates, warm-up steps and epochs. Of the fine-tuned models for each approach, the best-performing models were selected based on their metric scores, displayed in section 5. Initially, the dataset was shuffled with a fixed seed size of 42 and then used for fine-tuning and performance evaluation with a train-test split of 80 to 20 without additional randomization to ensure the reproducibility of the results. Since research and available methods for Text Simplification are limited but the task has significant overlaps with Machine Translation and Text Summarization, knowledge from these tasks was leveraged to build and test the proposed methods for automatic German Text Simplification. Benchmarking was performed on the Pegasus AI High-Performance Computing Cluster at DFKI, using 4 NVIDIA A100-PCIE-40GB GPUs, 2 CPUs per GPU, and 50GB of working memory. All models were fine-tuned within around 6 to 8 hours. They were fine-tuned with a defined maximum sequence length of 1024 tokens to process the given documents based on their found length and take into account the context of the full documents during their learned and applied translations to GE2R. As this was a computationally intensive approach for a complex task, we weighed costs of time and energy consumption against underlying model complexity and pre-training. Considering this, the following models were utilized.

4.1 Models and Tasks

The employed **mBART model for Summarization** is a fine-tuned checkpoint of the facebook/mbart-large-cc25 model (Liu et al., 2020) on the Multilingual Large Summarization dataset on several different languages, with a focus on Czech texts, to produce multilingual summaries and was obtained from huggingface.com[5]. It was provided by researchers at the Artificial Intelligence Center of the Czech Technical University in Prague[6]. This model can produce high-quality multi-sentence summaries in eight languages, including German, as demonstrated by its strong ROUGE benchmark scores on German texts.

The utilized **mBART model for Translation**[7] was fine-tuned for multilingual machine translation by Tang et al.. It supports a direct translation between any pair of the 50 languages (including German) and is a fine-tuned checkpoint of the original mBART-50 model, which exhibited an outstanding performance for translation tasks. The model was fine-tuned for ready-to-use multilingual translation tasks, supporting German as both the source and target language in translation pairs.

The **mt5 model for Summarization and Translation**[8] was pre-trained for both machine translation and text summarization in 101 different languages, including German. Although it produces state-of-the-art results for a wide variety of NLP tasks, it must be fine-tuned to be used for any downstream task using a task description in natural language. We chose the mt5-base model for its extensive multilingual pre-training and its aptitude for transfer learning based on its exceptional performance on many multilingual benchmarks (Xue et al., 2021).

4.2 Applied Metrics

Several readability metrics were used to assess the language complexity of the generated texts. The **Flesch Reading Ease** (FRE) metric (Flesch, 1948), commonly used to assess text complexity, provides interpretable numerical scores based on sentence and word length, influenced by syllables per word. Since the standard FRE penalizes characteristics typical of German, the adapted German FRE score from Toni Amstad's 1978 dissertation was used (Amstad, 1978). Scores range from 0 to 100, with 100 indicating the lowest complexity.

The **Wiener Sachtextformel** (WSTF) (Bamberger and Vanecek, 1984), designed to evaluate the readability of German factual texts, was used in this case. It combines features like word and sentence length, percentage of long words, and personal pronouns. Variant 1, a general-purpose metric for factual or educational texts, was applied. Scores typically range from 4.0 to 20.0, with higher scores indicating greater complexity. Scores below 4 are rarely achieved, as they denote texts that are exceptionally easy to read. Scores between 5 and 8 are considered accessible to the general public, including those with lower secondary education.

The **Coleman Liau Index** (CLI) (Coleman and Liau, 1975), a widely used readability metric for English, was applied here. It automatically analyzes digital content and assesses vocabulary complexity by counting the letters in words. Although designed for English, it serves as a comparative score to evaluate the complexity of source and reference texts for other languages. The score corresponds to US school grade levels (1-12), with higher scores indicating college or university literacy. Unlike most metrics, CLI does not consider syllables but uses the average number of letters per word and sentences per 100 words to compute the score.

Bilingual Evaluation Understudy (BLEU) (Papineni et al., 2002) and **Recall-Oriented Understudy for Gisting Evaluation** (ROUGE) (Lin, 2004) are widely used metrics for evaluating machine translation and summarization tasks. BLEU assesses the quality of translations by comparing n-grams in the generated output with reference translations with a focus on precision while penalizing a high output length. ROUGE, on the other hand, evaluates summarization by comparing n-grams or word sequences in the output texts with reference summaries, focusing on recall to evaluate content coverage. After the results were viewed, the ROUGE-L and ROUGE-LSUM values were found to be indistinguishable for all generated texts when applied at a document-level. Therefore, only ROUGE-L scores were provided.

The **Bilingual Annotation-based Content** (BLANC) score (Vasilyev et al., 2020) is a metric used for evaluating the quality of Text Summarization tasks. It assesses how well a generated summary preserves key content and relationships

[5] https://huggingface.co/ctu-aic/mbart25-multilingual-summarization-multilarge-cs

[6] https://www.aic.fel.cvut.cz/

[7] https://huggingface.co/facebook/mbart-large-50-many-to-many-mmt

[8] https://huggingface.co/google/mt5-base

from the original text without relying on human-written reference summaries. It introduces noise by randomly masking parts of the text, then uses a language model (Cloze task) to reconstruct the masked portions. The resulting score reflects how much of the original content is captured in the summary, typically ranging from 0.0 to 0.5, though it can reach 1.0. The score's interpretation depends on the task and data, so it's best used for comparison. Originally developed for English with a BERT model, it was adapted for German (Iskender et al., 2021) as follows using the bert-base-german-dbmdz-cased model, with a gap size of 2 and a maximum input token size of 256 for fine-tuning and inference.

4.3 Human Evaluation

Ten article pairs were selected, (see Appendix 8) consisting of a sample article in RG from the test set and a corresponding GE2R version generated using the fine-tuned mBART model with a Translation task preset. Texts with ROUGE-L scores above the median of the best-performing approach were chosen to showcase high-quality simplified texts, with a few containing common model errors to assess their impact on perceived quality and readability. The survey included questions on participants' education, reading level, preference, and impairment, along with six questions evaluating text quality, covering intent, content representation, logical consistency, grammar, reading ease, and use of GE2R in the generated articles. A total of 21 German native speakers with higher education participated. All participants submitted their responses regarding the featured article pairs, electively they could choose to answer the organizational parts of the survey in either RG or GE2R. This was done to ensure that the target group could be included and follow the given instructions and questions. After completion of the survey, we found that 14 participants completed the RG version, and 7 participants completed the GE2R version.

Of the RG participants, 35.7% preferred highly complex texts, 57.1% preferred medium-length, medium-complexity texts in everyday language, and 7.1% preferred short, low-complexity texts. For GE2R participants, 14.3% preferred long and complex texts, while 85.7% favored medium-length, medium-complexity texts. Fewer GE2R participants than RG participants expressed a preference for long texts if shorter, less complex alternatives with the same content were available.

Only 7.1% (one person) among the RG participants reported having a different reading comprehension. Explicit examples were given to indicate which group of people this may include, namely autistic people, people with ADHD, dyslexia, and hyperlexia. Notably though, 57.1% of the GE2R participants (four people) reported having different reading comprehension. When asked if they had experienced reading difficulties, none of the RG participants reported this, while 57.1% of the GE2R participants confirmed having such difficulties. Based on these observed differences in reading abilities and preferences, we separated the participants into two groups and present our findings as a comparison between the two groups.

5 Results and Discussion

Using the EasyGerman dataset, all models demonstrated the ability to improve their performance during fine-tuning and effectively applied their learned behavior during inference. Even the mBART summarization model, whose performance was most fluctuant during fine-tuning, improved over time. Detailed progressions of different metrics during fine-tuning are provided in Appendix 8.11. The performance comparison of all approaches, evaluated by automatic metrics, is shown in Table 6. It shows that the results of the generated articles using the test set at the end of fine-tuning in comparison to the generated articles using the validation set for inference. Since the distributed metric scores are non-uniform, the computed median scores are given as the scores for each metric. As readability scores are a comparative measure, the scores of the generated texts were compared to further computed readability scores of the reference texts in GE2R and the input texts in RG as seen in Table 7. The notable difference in scores between the input texts in RG and the generated texts in GE2R indicated a successful text simplification, showing a significantly lower linguistic complexity of the generated texts when compared to the reference texts in RG. The readability scores of the generated text also closely matched those of the professionally translated GE2R reference texts, suggesting a similarity of the generated texts to the references in terms of readability and simplicity. The near-identical readability scores for fine-tuning and inference demonstrate that the LLMs successfully learned and applied German simplification patterns, producing texts in simplified German with characteristics of GE2R.

Table 6: Metric Scores of Approaches for Fine-tuning and Inference

	mBART-sum		mBART-tr		mt5-sum		mt5-tr	
Metric	Fine-tuning	Inference	Fine-tuning	Inference	Fine-tuning	Inference	Fine-tuning	Inference
WSTF	4.4	4.5	*3.7*	3.9	*3.3*	**3.4**	*3.3*	**3.5**
FRE	*78*	79	*78*	**76**	79	78	76-77	49.5
CLI	15.52	13.84	*13.07*	**13.75**	12.01	12.69	12.08	12.67
TLiW*	91	105	157	156.5	*159.5*	**157**	152	**157**
BLANC	0.02	0.02	*0.06*	**0.06**	0.05	0.05	0.05	**0.06**
BLEU	0.02	0.03	*0.11*	0.10	0.10	**0.11**	0.11	0.10
BLEU P-1	0.51	0.50	*0.54*	**0.57**	0.51	0.54	0.50	0.53
BLEU P-2	0.17	0.17	*0.22*	**0.23**	0.21	0.22	0.21	0.23
BLEU P-3	0.06	0.06	*0.10*	**0.11**	*0.10*	**0.11**	*0.10*	**0.11**
BLEU P-4	0.02	0.02	0.04	0.05	*0.05*	**0.06**	*0.05*	0.04
ROUGE-1	0.31	0.31	*0.48*	**0.48**	0.44	0.44	0.44	0.44
ROUGE-2	0.11	0.11	*0.20*	**0.20**	0.18	0.18	0.18	0.19
ROUGE-L	0.17	0.17	*0.27*	**0.26**	0.26	**0.26**	*0.27*	**0.26**

* Text Length in words

All approaches produced low BLANC scores for both fine-tuning and inference, with the mBART Summarization model generating the lowest scores. This indicates that the generated texts lacked sufficient summarization qualities, confirming that transfer learning was successfully applied for text simplification, not summarization. Although BLEU scores were also low for all approaches, their performance in the context of automatic German text simplification aligned with scores reported in related work on generative German Text Simplification approaches (Rios et al., 2021) (Ebling et al., 2022). BLEU Precision-n-gram scores were higher for shorter n-grams and decreased steeply for longer ones. The mBART Translation approach generated the highest BLEU-Precision-n-gram scores, followed by mt5 Summarization, mt5 Translation, and mBART Summarization. For ROUGE scores, including ROUGE-L, the mBART Translation approach outperformed all others in both fine-tuning and inference, while the mt5 approaches showed nearly identical performance. The mBART Summarization approach yielded the lowest ROUGE and BLEU scores.

The ROUGE and BLEU metrics were most indicative of perceivable text quality, with higher ROUGE than BLEU scores suggesting that the models prioritized recall over precision. This indicates that the models focused on condensing or paraphrasing information, but were less precise in matching the reference language. The moderately high scores for smaller n-grams suggest significant lexical overlap with the references, though precision declined for longer sequences, potentially affecting text structure and comprehensiveness. Overall, the mBART Translation approach performed the best, as indicated by its high ROUGE and BLEU scores.

The texts generated by the mBART translation approach were well received overall by survey participants. Depictions of these responses for each assessed property can be seen in Appendix 8.12. Both groups generally agreed that the intent of the source texts was well represented in the generated texts. Overall, the majority of the RG group agreed that

Table 7: Median Metric Scores of References and Inputs for Fine-tuning and Inference.

	Fine-tuning		Inference	
Metric	**References**	**Inputs**	**References**	**Inputs**
WSTF	3.5	6.4	3.5	6.5
FRE	76-77	51	76-77	49.5
CLI	12.93	17.81	13.25	17.90
TLiW*	181	–	202.5	–
BLANC	0.06	–	0.06	–

* Text Length in words

the intent was very well represented for seven of the articles, with over 70 percent agreement for six of these articles and one article achieving only positive responses. Notably, the GE2R group showed even higher levels of agreement, with nine articles receiving over 70 percent positive responses and four articles achieving only positive responses. In contrast, the representation of content yielded more varied responses. Within the RG group, six articles received over 50 percent negative responses, with an overall tendency towards negative responses for this property. Although the GE2R group rated content representation more positively, responses were mixed, suggesting room for improvement.

Next, the quantity of grammatical errors was assessed. Responses showed that most texts contained few grammar errors which were likely not viewed as obtrusive by the majority of participants. Both groups showed similar proportions of participants reporting no errors. However, the GE2R group was more likely to perceive fewer errors overall. This could be attributed to differing reading proficiencies, which may influence error detection and perception. Errors in logical consistency were found to be more frequent and obtrusive to both participant groups than grammatical errors. The RG group, in particular, identified significant issues in this area, with the majority indicating a high number of logical inconsistencies in eight of the ten articles. Although the GE2R group reported fewer such errors, the presence of logical flaws was still noted. Regarding the readability of the generated texts, the GE2R group rated almost all articles as easier to read than the RG group. Except for article 10, which was unanimously deemed easy to read by both groups, the GE2R participants consistently reported notably higher readability ratings. Overall, the positive responses regarding the reading ease by both groups far outweighed the negative responses for most articles. This indicates that a degree of simplification could be achieved, which ensures that most articles are viewed as easy to read for both groups of participants and is well-received by members of its target group. Finally, the large majority of both groups agreed that all generated texts followed the principles of GE2R. The GE2R group unanimously recognized six articles as written in GE2R, while the RG group did so for five articles. This demonstrates the usefulness of the dataset for research on generating GE2R from regular German texts.

However, given that the target group of GE2R includes vulnerable individuals with impairments, further refinement in some identified areas and a more nuanced evaluation of responses to generated texts are likely needed before real-world deployment of this technology can be considered. Future research may also prioritize including a larger and more diverse group of participants with reading difficulties, representing different kinds of impairments and educational backgrounds. Nevertheless, based on the generally positive reception of the texts generated by the best-performing approach, the mBART translation model could serve as an assistive tool for human translators, helping to accelerate the simplification of news articles. This could improve the availability of informative content in GE2R.

6 Conclusion

This paper introduces a monolingual parallel dataset of German news article pairs published by the MDR, aligned at the document level, and containing simplified content written only in professionally translated GE2R. We present four approaches to assess the feasibility of generating GE2R using the EasyGerman dataset with Large Language Models (LLMs) for German Text Simplification using transfer learning. All approaches consistently achieved high scores across three readability metrics, indicating that the generated texts closely matched the readability and simplicity of the reference GE2R texts. These results demonstrate that LLMs can effectively learn and apply the rules to simplify German texts using the provided Dataset. Among the approaches, the fine-tuned mBART-50 model with a translation task preset produced the highest scores across BLEU and ROUGE metrics, showing strong capabilities for generating readable texts featuring complex document-level simplification properties.

A selection of texts generated by the best-performing model was well received in an online survey of 21 human evaluators. The survey assessed the generated articles on complex simplification qualities perceivable to humans such as intent, content representation, fluency, and logical consistency. The target group of GE2R, individuals with reading difficulties or disabilities, was included, enhancing the relevance of the findings. The majority of participants agreed that the generated texts accurately represented the source intent, were easy to read, and were written in GE2R. How-

ever, mixed responses regarding content representation were observed, potentially due to unfamiliarity with GE2R's inherent content reduction. Participants also noted some grammar errors and issues with logical consistency in the generated articles. These results suggest the need for further research to improve the quality of simplified texts and ensure the ethical use of automatic GE2R generation for real-world applications. Nonetheless, the feasibility of automatically generating GE2R from RG has been successfully demonstrated, providing valuable benchmarks for future research.

7 Limitations and Ethical Considerations

We would like to address several potential limitations and concerns. Firstly, the dataset comprises 1,916 sample pairs of documents, which may not constitute a sufficiently large dataset for training robust models. The dataset is constrained by the limited availability of parallel data in regular German and simplified GE2R. However, it is specifically constrained by the amount of parallel data published by the MDR until its finalization in April 2024 and made available on the described websites. Due to the design choice of including data published by a single source, overall quality and consistency in style and content impose limits on the ability of neural approaches to adapt to other document types or genres. This lack of diversity in data sources may negatively affect the generalizability of fine-tuned approaches for other domains or applications. Additionally, the reliance on news data may result in embedded biases, both in terms of content and style, that should be considered for future research.

Furthermore, no large language models with pretraining in German Text Simplification are publicly available. As a result, transfer learning was employed, aiming to leverage related knowledge to enhance performance. Given the task similarities, models pre-trained on machine translation and text summarization were utilized. However, the need for pre-training in German and the requirement for a large embedding dimension (around 1024) to process full documents significantly influenced the choice of models, thereby narrowing the selection of viable options.

We also discovered limitations in the useof text simplification metrics for the evaluation of results. Initially, we considered using D-SARI, the document-level adaption of SARI, a common text simplification metric. D-SARI first calculates SARI scores at the sentence level before computing the final score at the document level. However, due to the nature of GE2R, its application was found not feasible in this context. The significant length disparity between input and translated target texts, the loss of discourse structure during translation, and the impracticality of sentence mapping — neither manual nor automatic means were fund viable —rendered the computation of intermediate SARI and subsequent D-SARI scores ineffective.

The application of this research raises important ethical concerns since providing an accessible language for vulnerable groups is a sensitive area of application. Due to the nature of the target audience, careful attention to privacy, the accommodation of support needs of different groups, and an adequate representation is required. Efforts were made to involve organizations specializing in evaluating GE2R to include more representatives from the target group. Regrettably, logistical and practical challenges limited participation nonetheless. This underrepresentation may affect the model's fairness and efficacy in addressing the needs of the intended audience and should be considered in further research.

References

Toni Amstad. 1978. *Wie verständlich sind unsere Zeitungen?* Ph.D. thesis, Universität Zürich.

Dennis Aumiller and Michael Gertz. 2022. Klexikon: A German dataset for joint summarization and simplification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2693–2701, Marseille, France. European Language Resources Association.

Richard Bamberger and Erich Vanecek. 1984. Lesen-verstehen-lernen-schreiben: Die schwierigkeitsstufen von texten in deutscher sprache. *Jugend und Volk Verlagsgesellschaft*.

Alessia Battisti and Sarah Ebling. 2019. A corpus for automatic readability assessment and text simplification of german. *Preprint*, arXiv:1909.09067.

Alessia Battisti, Dominik Pfütze, Andreas Säuberli, Marek Kostrzewa, and Sarah Ebling. 2020. A corpus for automatic readability assessment and text simplification of German. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3302–3311, Marseille, France. European Language Resources Association.

Sofia Blinova, Xinyu Zhou, Martin Jaggi, Carsten Eickhoff, and Seyed Ali Bahrainian. 2023. SIMSUM:

Document-level text simplification via simultaneous summarization.

Bundesministerium der Justiz. 2002. Behindertengleichstellungsgesetz abschnitt 2 verpflichtung zur gleichstellung und barrierefreiheit. § 11 verständlichkeit und leichte sprache.

Meri Coleman and T. L. Liau. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60:283–284.

Sarah Ebling, Alessia Battisti, Marek Kostrzewa, Dominik Pfütze, Annette Rios, Andreas Säuberli, and Nicolas Spring. 2022. Automatic text simplification for german. *Frontiers in Communication*, 7.

R. F. Flesch. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–223.

Silvia Hansen-Schirra, Jean Nitzke, and Silke Gutermuth. 2021. An intralingual parallel corpus of translations into german easy language (geasy corpus): What sentence alignments can tell us about translation strategies in intralingual translation. In Vincent X. Wang, Lily Lim, and Defeng Li, editors, *New Perspectives on Corpus Translation Studies*, New Frontiers in Translation Studies, pages 281–298. Springer.

Renate Hauser, Jannis Vamvas, Sarah Ebling, and Martin Volk. 2022. A multilingual simplified language news corpus. In *Proceedings of the 2nd Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI) within the 13th Language Resources and Evaluation Conference*, pages 25–30, Marseille, France. European Language Resources Association.

Neslihan Iskender, Oleg V. Vasilyev, Tim Polzehl, John Bohannon, and Sebastian Möller. 2021. Towards human-free automatic quality evaluation of german summarization. *CoRR*, abs/2105.06027.

Sarah Jablotschkin, Elke Teich, and Heike Zinsmeister. 2024. DE-lite - a new corpus of easy German: Compilation, exploration, analysis. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 106–117, St. Julian's, Malta. Association for Computational Linguistics.

Sarah Jablotschkin and Heike Zinsmeister. 2020. Leiko: A corpus of easy-to-read german. In *Poster presented at the Computational Linguistics Poster Session in the course of the 42nd annual conference of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) in Hamburg. Download of Poster and corpus from: https://doi. org/10.5281/zenodo*, volume 3923917.

David Klaper, Sarah Ebling, and Martin Volk. 2013. Building a German/simple German parallel corpus for automatic text simplification. In *Proc. 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 11–19, Sofia, Bulgaria. Association for Computational Linguistics.

Daisy Lange. 2018. Comparing 'leichte sprache','einfache sprache'and 'leicht lesen': A corpus-based descriptive approach. In *Eds. SUSANNE J. JEKAT, and GARY MASSEY. Barrier-free communication: methods and products: proceedings of the 1st Swiss conference on barrier-free communication. Winterthur: ZHAW Digital Collection*, pages 75–91.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Margot Madina, Itziar Gonzalez-Dios, and Melanie Siegel. 2023. Easy-to-read in germany: A survey on its current state and available resources.

Babak Naderi, Salar Mohtaj, Kaspar Ensikat, and Sebastian Möller. 2019. Subjective assessment of text complexity: A dataset for german language. *arXiv preprint arXiv:1904.07733*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Annette Rios, Nicolas Spring, Tannon Kew, Marek Kostrzewa, Andreas Säuberli, Mathias Müller, and Sarah Ebling. 2021. A new dataset and efficient baselines for document-level text simplification in German. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 152–161, Online and in Dominican Republic. Association for Computational Linguistics.

Andreas Säuberli, Sarah Ebling, and Martin Volk. 2020. Benchmarking data-driven automatic text simplification for German. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*, pages 41–48, Marseille, France. European Language Resources Association.

Regina Stodden, Omar Momen, and Laura Kallmeyer. 2023. DEplain: A German parallel corpus with intralingual translations into plain language for sentence and document simplification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16441–16463, Toronto, Canada. Association for Computational Linguistics.

Renliang Sun, Hanqi Jin, and Xiaojun Wan. 2021. Document-level text simplification: Dataset, criteria and baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7997–8013, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2020. Multilingual translation with extensible multilingual pretraining and finetuning. *CoRR*, abs/2008.00401.

Vanessa Toborek, Moritz Busch, Malte Boßert, Christian Bauckhage, and Pascal Welke. 2022. A new aligned simple German corpus. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11393–11412, Toronto, Canada. Association for Computational Linguistics.

Oleg Vasilyev, Vedant Dharnidharka, and John Bohannon. 2020. Fill in the blanc: Human-free quality estimation of document summaries. *arXiv preprint arXiv:2002.09836*.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

8 Appendix

The source articles are part of the EasyGerman dataset and were used to generate the second article of each pair. The source articles are written in regular German and the generated articles are written in simplified German with the expectation of containing language written in Leichte Sprache (GE2R). They were produced by mBART translation model fine-tuned on the EasyGerman dataset. The following ten articles pairs were featured and evaluated in the given order.

8.1 Article Pair 1

8.1.1 Source Text

Erste Briefe im Weihnachtspostamt Himmelsberg angekommen | MDR.DE

Alljährlich schreiben tausende Kinder an den Weihnachtsmann - die Thüringer Zweigstelle ist in Himmelsberg. Im letzten Jahr sind rund 7.000 Briefe von nah und fern im Kyffhäuserkreis eingegangen. Auch in diesem Jahr wollen wieder neun Frauen bis Heiligabend jeden Brief beantworten.

Im einzigen Thüringer Weihnachtspostamt in Himmelsberg im Kyffhäuserkreis sind die ersten 30 Briefe angekommen. Wie Ramona Schattney vom Heimatverein MDR THÜRINGEN sagte, haben unter anderem Kinder aus Thailand, Taiwan und Deutschland geschrieben.

Der große Ansturm mit täglich bis zu 200 Briefen beginne meistens erst im Advent. Auch in diesem Jahr wollen wieder neun Frauen bis Heiligabend jeden Brief beantworten. Im Vorjahr waren mehr als 7.000 Briefe angekommen. Derzeit wird das Antwortschreiben in der Druckerei vorbereitet, tausende Briefmarken liegen ebenfalls bereit.

Weihnachtsmärchen wird schon geprobt

Kopfzerbrechen macht den Helfern des Weihnachtsmannes noch der traditionelle Tag der offenen Tür am ersten Adventssamstag. Derzeit hoffen die Ehrenamtler, dass das Backhaus wieder regulär öffnen darf. Das Märchen werde schon fleißig geprobt. Die Thüringer Helfer des Weihnachtsmannes treffen sich wie im Vorjahr im beheizten Dorfgemeinschaftshaus.

Wegen Corona musste das Weihnachtsmannbüro in den beiden Vorjahren geschlossen bleiben. Trotzdem waren insgesamt fast 14.000 Briefe an das Weihnachtsmannbüro in 99713 Himmelsberg gekommen.

8.1.2 Generated Text

Leichte Sprache: Im Weihnachts-Postamt in Himmelsberg sind die ersten Briefe angekommen | MDR.DE

Himmelsberg ist ein Orts-Teil von der Stadt Sondershausen. Das ist im Bundes-Land Thüringen. Dort gibt es ein besonderes Postamt. Die Menschen können einen Brief dorthin schicken. Sie schreiben den Brief an den Weihnachts-Mann. In dem Brief können sie zum Beispiel schreiben: Was sie sich zu Weihnachten wünschen. Dann schreibt ihnen der Weihnachts-Mann einen Brief zurück. Beim Weihnachts-Postamt in Himmelsberg sind jetzt die ersten Briefe angekommen. Ramona Schattney arbeitet beim Heimat-Verein MDR Thüringen. Sie hat gesagt: Viele Menschen schreiben Briefe an den Weihnachts-Mann. Die Menschen schreiben zum Beispiel Briefe aus verschiedenen Ländern. Zum Beispiel: • Aus dem Land Thailand, • aus dem Land Taiwan • und aus Deutschland. In Himmelsberg gibt es das Büro vom Weihnachts-Mann. Dort arbeiten die Mitarbeiter vom Weihnachts-Postamt bis zum Heiligabend. In diesem Jahr wollen wieder 9 Frauen alle Briefe zurück schreiben.

8.2 Article Pair 2

8.2.1 Source Text

DFB-Pokal: RB Leipzig besiegt Hoffenheim und steht im Viertelfinale | MDR.DE
Achtelfinale

Heißer Sieg bei eisigen Temperaturen: RB Leipzig hat am Mittwochabend die TSG Hoffenheim mit 3:1 2:0 bezwungen und sich damit ins Pokal-Viertelfinale vorgearbeitet. Es war ein verdienter Erfolg, der nur in einer kurzen Phase gefährdet schien. Der Jubel über Tore und Sieg war wegen eines Notfalleinsatzes nur verhalten.

Die Erfolgsserie geht auch im Pokal weiter: RB Leipzig hat Hoffenheim 3:1 bezwungen und

ist damit nun seit 17 Pflichtspielen in Folge ungeschlagen. Überschattet wurde die Partie allerdings von einem Notfall vor dem Anpfiff: Während es am Himmel blitzte und donnerte, musste am Einlass ein Fan reanimiert werden. Daraufhin verzichteten die Leipziger Anhänger weitgehend auf eine lautstarke Unterstützung ihres Teams.

Blitz, Donner, frühe Führung

Der Partie vor knapp 35.000 Zuschauern ging ein eisiges Gewitter voraus. Nach dem Schließen der Schleußen stellten die Leipziger die klar bessere Mannschaft, hatten mehr Ballbesitz, zeigten mehr Initiatve, waren gefährlicher. Ab und an gingen sie auch vom Gas - aber in genau so einer Phase fiel das 1:0 durch Emil Forsberg, der auf der linken Seite den Ball ungestört runterpflücken, querlaufen und dann aus 17 m einschießen konnte. Das 2:0 dagegen fiel in einer Drangphase bzw. im Nachsetzen: Konrad Laimer traf mit einem Linksschuss aus 15 m ins lange linke Eck 42. Hoffenheim kam nur selten vor das RB-Tor, war aber meist nur dann gefährlich, wenn die Leipziger sich Stockfehler oder Missverständnisse erlaubten.

RBL-Wackelphase nach dem Anschluss

Das große Aufbäumen der Gäste ließ auch nach der Pause auf sich warten. Statt-dessen war RB am Drücker, blieb beim letzten Pass jedoch zu ungenau oder zu spät, sodass die Abseitsfahne oft hochschnellte. Wie aus dem Nichts kam dann die TSG zum Anschluss: Der eingewechselte Kasper Dollberg veredelte eine genaue Angelino-Flanke zum 1:2. Wie zuletzt gegen Stuttgart wackelte RB nach dem Gegentor, stand mächtig unter Stress. Diesmal musste aber nicht bis in die Nachspielzeit gezittert werden, denn ein Konter in der 84. Minute machte den Deckel auf die Partie. Timo Werner traf nach starker Vorarbeit von Dominik Szoboszlai und dem eingewechselten Yussuf Poulsen.

8.2.2 Generated Text

Leichte Sprache: RB Leipzig hat am Mittwoch im DFB-Pokal gegen Hoffenheim gewonnen | MDR.DE

Hoffenheim ist eine Stadt im Bundes-Land Sachsen. Von dort kommt der Fußball-Verein RB Leipzig. Er hatte am Mittwoch-Abend ein Fußball-Spiel gegen den Fußball-Verein TSG Hoffenheim. Und bei diesem Fußball-Spiel hat RB Leipzig gewonnen. Die Fußball-Spieler von TSG Hoffenheim haben gut gespielt. Deshalb haben sie das Fußball-Spiel gewonnen. Jetzt ist RB Leipzig im Viertel-Finale vom DFB-Pokal. Wenn der Fußball-Verein RB Leipzig das Fußball-Spiel gewinnt,dann ist der Verein im Viertel-Finalevon dem DFB-Pokal.

8.3 Article Pair 3

8.3.1 Source Text

Nach Waldbränden: Sächsisches Kabinett setzt Expertenkommission ein | MDR.DE

Die sächsische Regierung hat am Dienstag die Einsetzung einer unabhängigen Expertenkommission beschlossen, informiert die Sächsiche Staatskanzlei. Man wolle dabei Konsequenzen aus den verheerenden Waldbränden in der Gohrischheide, in Arzberg sowie im Nationalpark Sächsische Schweiz ziehen. In dem Gremium werden Personen aus unterschiedlichen Fachbereichen diskutieren.

Nach den Waldbränden mit Millionenschäden hatte die sächsische Regierung Schutz-konzepte angekündigt. Nun hat sie dafür ein entsprechendes Gremium eingesetzt. Ziel der Kommission soll es sein, die Waldbrände objektiv auszuwerten. Dadurch sollen Maßnahmen für eine bessere Prävention und Bekämpfung von Waldbränden entwickelt werden.

Laut Staatskanzleichef Oliver Schenk gilt es nach dem Abschluss der Löscharbeiten nun Lehren für die Zukunft zu ziehen. Dafür soll die unabhängige Expertenkommission Vorschläge erarbeiten.

Den Vorsitz der Kommission soll Hermann Schröder übernehmen, der als langjähriger Leiter der Abteilung "Bevölkerungsschutz und Krisenmanagementïn Baden-Württemberg ein ausgewiesener Fachmann in dem Bereich sei, heißt es weiter.

Bisherige Schutzkonzepte sollen mit einfließen

Die Expertenkommission will auch auf bereits gewonnene Erkenntnisse etwa aus der landesweiten SStrategischen Waldbrandschutzkonzeption"des Innenministeriums oder dem Waldbrandschutzkonzept des Umweltministeriums für die Nationalparkregion zurückgreifen. Zudem könnten Erkenntnisse durch das Brandgeschehen von tschechischer Seite mit berücksichtigt werden.

8.3.2 Generated Text

Leichte Sprache: In Sachsen gibt es eine Fach-Kommission nach Waldbränden | MDR.DE

Die Regierung von dem Bundes-Land Sachsen hat entschieden: Es soll eine Fach-Kommission geben. Sie soll über die Waldbrände sprechen. Die Fach-Kommission soll zum Beispiel darüber sprechen: • Welche Sachen die Menschen in den Wäldern machen sollen. • Welche Regeln gelten: Damit die Waldbrände besser verhindert werden können. • Und damit die Waldbrände besser gelöscht werden können. Oliver Schenk ist der Chef von der Staats-Kanzlei. Er hat gesagt: Die Fach-Kommission soll lernen: Was die Menschen bei den Waldbränden lernen können. Das werden verschiedene Menschen machen. Zum Beispiel: • Menschen aus verschiedenen Fach-Bereichen. • Und Menschen mit verschiedenen Behinderungen. Die Menschen wollen in der Fach-Kommission auch darüber sprechen: Was die Menschen bei den Waldbränden lernen können. Zum Beispiel: • Über die vielen Waldbrände in den Bundes-Ländern: • Sachsen, • Sachsen-Anhalt • und Thüringen. Und was die Menschen bei den Waldbränden machen können.

8.4 Article Pair 4

8.4.1 Source Text

Magische Lichterwelten im Bergzoo Halle bis 2. März verlängert | MDR.DE

Bisher haben mehr als 80.000 Menschen die Magischen Lichterwelten im Bergzoo in Halle Saale gesehen. Am Ende soll wieder die Marke von 100.000 geknackt werden. Die Veranstalter rechnen noch einmal mit großem Andrang. Sie haben die Lichterwelten deswegen um eine Woche bis zum 2. März verlängert.

Die magischen Lichterwelten im Bergzoo Halle werden verlängert. Das hat Veranstaltungschef Tom Bernheim MDR SACHSEN-ANHALT gesagt. 80.000 Besucher hätten die Lichterwelten seit der Eröffnung vor zwei Monaten bereits gesehen. Dass die 100.000 angepeilten Besucher erreicht werden, gilt als sicher.

Ende erst am 2. März großer Andrang erwartet

Erfahrungsgemäß gibt es Bernheim zufolge zum Ende hin einen großen Andrang. Um dem gerecht zu werden, wird die Schau um genau eine Woche verlängert und endet nicht am 24. Februar sondern erst am 2. März. Die magischen Lichterwelten zeigen in diesem Jahr Szenen und Figuren aus Tausendundeiner Nacht. MDR SACHSEN-ANHALT ist Medienpartner der magischen Lichterwelten.

8.4.2 Generated Text

Leichte Sprache: Im Zoo von der Stadt Halle gibt es eine besondere Lichter-Welt | MDR.DE

Die Stadt Halle ist in dem Bundes-Land Sachsen-Anhalt. Dort gibt es einen Zoo. Der Zoo heißt: Bergzoo Halle. Dort werden in diesem Jahr besondere Lichter-Welten gezeigt. Sie dauern eine Woche. Bis zum 2. März können die Menschen wieder zu den Lichter-Welten gehen. Der Chef von den Lichter-Welten heißt: Tom Bernheim. Er hat gesagt: Seit der Eröffnung von den Lichter-Weltenhaben schon mehr als 80 tausend Menschendie Lichter-Welten gesehen. Und die Besucher werden sich die Lichter-Welten schon angeschaut. Das bedeutet: Sie werden die Lichter-Welten schon wieder ansehen. Tom Bernheim hat auch gesagt: Die Lichter-Welten werden eine Woche länger gemacht. Das bedeutet: Die Menschen können noch bis zum 2. März zur Lichter-Welten gehen.

8.5 Article Pair 5

8.5.1 Source Text

Neue Lithium-Fabrik für mehr als 100 Millionen Euro in Bitterfeld geplant | MDR.DE

Im Bitterfeld-Wolfen bahnt sich eine wirtschaftliche Großansiedlung an: Im Chemie-park soll eine Lithium-Raffinerie entstehen. Ein amerikanisch-holländisches Unternehmen will mehr als 100 Millionen Euro in die Anlage investieren. 2023 soll die Produktion beginnen. Die Stadt Bitterfeld hofft, dass die Fabrik weitere Unternehmen anlockt.

Im Chemiepark Bitterfeld entsteht ein großes Werk zur Aufbereitung von Lithium. Der amerikanisch-holländische Metall-Konzern Advanced Metallurgical AMG kündigte an, 120 Millionen Dollar, umgerechnet mehr als 103 Millionen Euro, zu investieren. Das Lithium soll laut Investor eine so gute Qualität haben, dass es für die Produktion von Batterien infrage kommt. Das Grundstück für die Raffinerie hatte der Konzern bereits im Frühjahr gekauft.

Fabrik soll 20.000 Tonnen Lithium fördern
Lithium ist ein begehrter, weil knapper Rohstoff. Er wird unter anderem bei der Produktion von Batterien für E-Autos gebraucht. Für die Fabrik in Bitterfeld wird zunächst mit 20.000 Tonnen pro Jahr geplant. Tritt die erwartete hohe Nachfrage ein, könne man die Menge auch verfünffachen, hieß es von AMG.

Der Bau der Anlage soll schon im ersten Quartal 2022 beginnen. Nach Unternehmensangaben laufen gerade Gespräche mit dem Bauordnungsamt. Produktionsstart soll dann im Herbst 2023 sein. Bitterfeld-Wolfens Oberbürgermeister Armin Schenk hofft auf die Schaffung zukunftsorientierter Arbeitsplätze in Bitterfeld-Wolfen und darauf, dass die AMG-Ansiedlung weitere Unternehmen in seine Stadt lockt.

8.5.2 Generated Text

Leichte Sprache: In Bitterfeld soll es eine neue Fabrik für Lithium geben | MDR.DE

Bitterfeld ist eine Stadt im Bundes-Land Sachsen-Anhalt. Dort will eine Firma ein neues Werk bauen. Die Firma heißt: AMG Bitterfeld. Das ist eine große Firma aus dem Land: USA. In dem Werk sollen Batterien für E-Autos hergestellt werden. Die Firma hat gesagt: Wir brauchen das Lithium. Deshalb soll es eine neue Fabrik im Jahr 2023 geben. Armin Schenk ist der Ober-Bürgermeister von Bitterfeld-Wolfen. Er wünscht sich: Dass mehr Firmen für die Fabrik kommen. Und dass sie für die Herstellung von den Batterien bald Arbeit haben.

8.6 Article Pair 6

8.6.1 Source Text

Verfassungsschutz: AfD Sachsen-Anhalt gesichert rechtsextremistisch | MDR.DE

Der Verfassungsschutz hat den Landesverband Sachsen-Anhalt der AfD als gesichert rechtsextremistische Bestrebung eingestuft.

Der Verfassungsschutz von Sachsen-Anhalt hat den AfD-Landesverband als gesichert rechtsextremistische Bestrebung eingestuft. Das hat der Leiter der Behörde, Jochen Hollmann, MDR SACHSEN-ANHALT mitgeteilt.

AfD Sachsen-Anhalt seit 2021 Verdachtsfall

Seit Beginn der Beobachtung hätten sich die tatsächlichen Anhaltspunkte für das Vorliegen einer Bestrebung gegen die freiheitliche demokratische Grundordnung sowohl in qualitativer als auch in quantitativer Hinsicht weiter verdichtet. Die nun erfolgte Einstufung als gesichert rechtsextremistische Bestrebung sei die rechtliche Konsequenz.

Hollmann teilte mit, dass die AfD Sachsen-Anhalt seit der Einstufung als Verdachtsfall im Januar 2021 intensiv geprüft worden sei. Man habe sich bei der Bewertung streng an dem gesetzlichen Auftrag des Verfassungsschutzgesetzes orientiert.

Verfassungsschutz: AfD hat sich seit Corona-Pandemie radikalisiert

Hollmann sagte: "Das Ergebnis ist eindeutig: Der Landesverband vertritt nicht nur weiterhin verfassungsfeindliche Positionen, die zur Einstufung als Verdachtsfall geführt hatten, sondern hat sich vielmehr seit der Corona-Pandemie derart radikalisiert, dass eine systematische Beobachtung unter Einsatz nachrichtendienstlicher Mittel gerechtfertigt ist."

Die Verfassungsschutzabteilung sei gesetzlich

dazu verpflichtet, die Einstufung als gesichertes Beobachtungsobjekt im Phänomenbereich Rechtsextremismus vorzunehmen. Es gebe eine "Vielzahl von Aussagen des Landesverbandes und zahlreicher Funktions- und Mandatsträger der AfD Sachsen-Anhalt, welche die Bestrebungen dieser Partei gegen die freiheitliche demokratische Grundordnung belegen", so Hollmann.

Der Verfassungsschutz komme seinem ëbenfalls gesetzlich definierten Informationsauftrag als Frühwarnsystem der Demokratie"nach, indem er die Entscheidung öffentlich bekannt gibt.

AfD strebt laut Verfassungsschutz gegen freiheitliche demokratische Grundordnung

Die AfD Sachsen-Anhalt verstößt nach Angaben des Verfassungsschutzes gegen das Prinzip der Menschenwürde, gegen das Demokratieprinzip und das Rechtsstaatsprinzip.

MDR-Informationen zufolge umfasst die Beweissammlung mehr als einhundert Seiten. Demnach wurden vor allem öffentliche Quellen wie YouTube-Mitschnitte von Parteiveranstaltungen, Wahlkampfreden und Interviews ausgewertet.

Verstoß gegen Prinzip der Menschenwürde

Äußerungen der AfD Sachsen-Anhalt, die mit der Garantie der Menschenwürde unvereinbar sind, bezeichnete der Verfassungsschutz als "besonders relevant". Der Verfassungsschutz teilte mit, dass führende Vertreter der Partei Menschengruppen abwerten, indem sie Migranten zum Beispiel als Ïnvasoren", Ëindringlingeöder "kulturfremde Versorgungsmigranten"diffamieren oder deutsche Staatsbürgerinnen und Staatsbürger mit Migrationshintergrund als "Passdeutsche"bezeichnen.

Hollmann zufolge hat der Verfassungsschutz zahlreiche muslimfeindliche, rassistische und antisemitische Aussagen von Funktions- und Mandatsträgern des Landesverbandes ausgewertet. Sie belegten, dass die Partei ein ëthnokulturell homogenes Staatsvolkänstrebe und die Ausgrenzung von Menschen aufgrund ihrer Herkunft oder Religion fordert.

Verstoß gegen das Demokratieprinzip

Außerdem strebe die Partei die Abschaffung der parlamentarischen Demokratie an. Die AfD Sachsen-Anhalt versuche, "das demokratische System der Bundesrepublik Deutschland sowie seine Institutionen und deren Vertreter verächtlich zu machen."Der Verfassungsschutz spricht von einer Delegitimierungskampagne.

Insbesondere während der Corona-Pandemie habe der Landesverband die Bundesrepublik immer wieder mit autokratischen oder gar totalitären Regimen sowie die Coronamaßnahmen mit der Judenverfolgung im Dritten Reich gleichgesetzt.

Auch dabei wurden antisemitisch geprägte Begriffe herangezogen, etwa die Verschwörungstheorie des "Great Reset", nach der eine globale Elite die Pandemie instrumentalisiert oder geplant habe, eine von ihr gesteuerte neue Weltordnung zu errichten.

Verstoß gegen das Rechtsstaatsprinzip

Gegen das Rechtsstaatsprinzip würden Äußerungen der AfD Sachsen-Anhalt verstoßen, in denen der Verfassungsschutz das Bestreben erkenne, ganze soziale Gruppen zu entrechten und einer faktischen Willkürherrschaft zu unterwerfen.

Die AfD Sachsen-Anhalt sowie ihre Teil- und Unterorganisationen würden die Ideologie des Ethnopluralismus vertreten. Das deute darauf hin, dass sich die Ziele der Partei nicht ohne die Verletzung rechtsstaatlicher Grundsätze umsetzen lassen.

AfD in Thüringen bereits als "gesichert rechtsextremistisch"

In Thüringen war die AfD vor zwei Jahren vom Verfassungsschutz als gesichert rechtsextremistisch eingestuft worden. In Sachsen-Anhalt und Brandenburg gilt die Einstufung für die AfD-Jugendorganisation Junge Alternative.

AfD: Entscheidung politisch motiviert

AfD-Co-Fraktionschef Oliver Kirchner sagte: Ës interessiert mich nicht, was der Verfassungsschutz behauptet."Laut Kirchner ist die Einschätzung politisch motiviert. Das Innenministerium werde von der CDU geführt, an der die AfD laut einer

kürzlich erschienenen Wählerumfrage vorbeigezogen ist, so der Co-Chef.

Landespolitiker nicht überrascht

Eva von Angern, die Fraktionschefin der Linken im Landtag von Sachsen-Anhalt, sagte wie mehrere Landespolitiker, sie sei nicht überrascht von der Entscheidung. "Überall in Sachsen-Anhalt erleben wir regelmäßig, wie diese Partei Grund- und Menschenrechte mit Füßen tritt."

Der innenpolitische Sprecher der SPD-Landtagsfraktion, Rüdiger Erben, sagte der Nachrichtenagentur dpa, er könne die Entscheidung nachvollziehen. Das Problem des Rechtsextremismus sei damit aber nicht gelöst.

Die Landesvorsitzende der Grünen, Madeleine Linke, schrieb auf X ehemals Twitter : "Nun ist bestätigt, was wir alle schon lange wussten: Die noAfD ist eine rechtsextreme Partei. Denn diese Partei geht von der Ungleichwertigkeit von Menschen aus, verharmlost den Nationalsozialismus hat eine starke Affinität zu diktatorischen Regierungsformen...".

8.6.2 Generated Text

Leichte Sprache: Der Verfassungs-Schutz von Sachsen-Anhalt hat die AfD als rechts-extremistisch ein-gestuft | MDR.DE

Im Bundes-Land Sachsen-Anhalt gibt es den Verfassungs-Schutz. Er kümmert sich um die Verfassung von Sachsen-Anhalt. Und er hat jetzt die AfD-Landes-Verband als gesichert rechts-extremistisch ein-gestuft. Das bedeutet: Der Verfassungs-Schutz hat jetzt gesagt: Die AfD ist eine rechts-extremistisch-extreme Gruppe. Das ist auch im Bundes-Land Sachsen-Anhalt so. Der Verfassungs-Schutz hat auch gesagt: Die AfD ist eine rechts-extreme Gruppe. Es gibt viele Hinweise dafür: Dass die AfD bestimmte Regeln nicht be-achtet. Der Verfassungs-Schutz hat zum Beispiel diese Regeln geprüft: • Wenn die AfD rechts-extrem ist: Dann ist das eine Straftat. • Wenn die AfD rechts-extrem ist: Dann ist das eine Straftat. • Und wenn die AfD rechts-extrem ist: Dann ist das eine Straftat. Der Verfassungs-Schutz hat dazu gesagt: Die AfD-Landes-Verband ist als gesichert rechts-extremistisch ein-gestuft. Das bedeutet: Der Verfassungs-Schutz hat als gesichert rechts-extremistisch ein-gestuft. Das hat Jochen Hollmann der Presse gesagt. Er ist der Chef von dem Verfassungs-Schutz.

8.7 Article Pair 7

8.7.1 Source Text

Annaberg-Buchholz macht sich fit für die Forschung zum autonomen Bahnfahren | MDR.DE

Es ist nicht Stanford oder Cambridge - noch nicht einmal Chemnitz: In Annaberg-Buchholz werden die Weichen für das Bahnfahren ohne Lokführer gestellt. Möglich machen das eine Teststrecke, Forscherinnen und Forscher und der SSmart Rail Connectivity Campus", der im Unteren Bahnhof der Erzgebirgsstadt entsteht.

Spitzenforschung unter dem Schwibbogen, so nennt der Rektor der Technischen Universität Chemnitz, Gerd Strohmeier, die Arbeit im SSmart Rail Connectivity CampusSSRCC im Unteren Bahnhof von Annaberg-Buchholz. Zur Eröffnung hatte er schon ganz spontan die Erzgebirgsstadt zum Universitätsstandort erklärt.

"Wir machen das auf einem sehr, sehr hohen Niveau hier in Annaberg-Buchholz, unter weltweit einzigartigen Rahmenbedingungen."Man könne es nicht genug betonen: "Wir forschen hier nicht im Stanford, Cambridge oder auch in Chemnitz. Wir machen das in Annaberg-Buchholz."

Sehr gute Bedingungen für Theorie und Praxis

Kern der Forschung seien neben dem Campus auch die praxisorientierten Rahmenbedingungen, so Strohmeier. "Wir haben hier eine mehr als 20 Kilometer lange Teststrecke zwischen Annaberg-Buchholz und Schwarzenberg, die mit einem 5G-Testfeld ausgerüstet wurde."

Das ermögliche zum Beispiel die Fernsteuerung realer Züge. "Das Ziel muss weiterhin sein, das automatisierte Bahnfahren weiter voranzutreiben."Nachhaltigkeit sei aber auch ein wichtiges Thema in Bezug auf alternative Antriebe und

Gewichtsersparnis bei den Zügen.

Kopfbau mit Köpfchen

Im neu eingeweihten nördlichen Kopfbau werden auch die forschenden Köpfe in Sachen Bahnforschung ihren Platz finden. Auf drei Etagen sind Büros, Labore und Server-Räume entstanden. Mit der Frauscher Sensortechnik GmbH hat auch ein Projektpartner aus der Wirtschaft Platz gefunden.

Die Ziele des SRCC sind hoch gesteckt: Sie reichen von der digitalen Vernetzung und Kommunikation im Schienenverkehr über das automatisierte Fahren auf der Schiene bis hin zur Integration des Bahnverkehrs in Mobilitätsangebote mit anderen Verkehrsmitteln.

Ziel: Mobilität im ländlichen Raum verbessern

Auf dem Campus sollen bahntechnische Erfindungen entwickelt, getestet und zur Marktreife gebracht werden, sagt der technische Leiter des SRCC, Sören Claus. "Wir hatten ja erst im November eine Fernsteuerfahrt, bei der ein Zug in Schlettau von Braunschweig aus gesteuert wurde über 5G."Man arbeite aber auch an Themen wie "Mobilitätsketten im ländlichen Raum", um das Leben dort attraktiver zu machen.

SSignaltechnik, Sensorik und 5G kommen dazu, wobei dieser Übertragungsstandard eine sehr große Rolle spielt", fügt Claus hinzu.

Wissenschaftsminister Gemkow hat auch Erwartungen an den Bund

Sachsens Wissenschaftsminister Sebastian Gemkow CDU fordert bei aller Freude auch eine weitere Beteiligung des Bundes. "Der Freistaat hat das Projekt schon in der Vergangenheit unterstützt und wird das auch weiterhin tun. Wir erwarten aber auch vom Bund, dass er das Projekt finanziell weiter unterstützt.SSachsen habe im Doppelhaushalt bereits die entsprechenden Mittel eingestellt. Genaue Summen könne er jedoch wegen der dynamischen Preisentwicklung am Bau nicht nennen, sagte Gemkow. Im nächsten Schritt soll zwischen den beiden Kopfbauten des Bahnhofs eine Forschungshalle gebaut werden.

8.7.2 Generated Text

Leichte Sprache: In Annaberg-Buchholz werden viele Sachen für die Autonome Bahn gemacht | MDR.DE

Die Stadt Annaberg-Buchholz ist in dem Bundes-Land Sachsen. Dort gibt es die Firma: Smart Rail Connectivity Campus. Die Abkürzung dafür ist: SRCC. In der Stadt Annaberg-Buchholz gibt es einen neuen Campus für Bahn-Forschung. Dort wollen einige Forscher heraus-finden: Wie die Züge ohne Lok-Führer fahren können. In dem Campus werden viele Sachen für die Autonome Bahn gemacht. Zum Beispiel: • Eine Test-Strecke zwischen Annaberg-Buchholz und Schwarzenberg. Sie ist 20 Kilometer lang. In der Test-Strecke sind auch 5G-Test-Flächen ausgerüstet. So können die Forscher zum Beispiel heraus-finden: • Ob die Züge eine Fern-Steuer-Funktion haben. • Oder ob sie Züge ohne Lok-Führer fahren können. In dem Campus sollen noch mehr Sachen gemacht werden. Zum Beispiel: • Büros, • Labore • und Server-Räume.

8.8 Article Pair 8

8.8.1 Source Text

Hochwasser in Sachsen-Anhalt: Hohe Pegel an Ohre, Schwarzer Elster und im Harz | MDR.DE

Die Pegelstände in Sachsen-Anhalt sind in den vergangenen Tagen vielerorts angestiegen. In Wolmirstedt, Löben und Ilsenburg wurden auch erste Alarmstufen erreicht. Grund sind die vielen Niederschläge und die durch die milden Temperaturen einsetzenden Niederschläge. Dramatisch sei die Situation aber nicht, sagt ein Experte.

Viele Niederschläge und das milde Wetter und damit verbunden die Schneeschmelze haben vielerorts in Sachsen-Anhalt die Pegelstände ansteigen lassen. In Löben an der Schwarzen Elster und in Ilsenburg im Harz lagen die Pegel im Bereich von Alarmstufe 1. An der Ohre in Wolmirstedt wurde Alarmstufe 2 von 4 erreicht.

Ohre: Wachsam sein, aber keine dramatische Situation

Der Direktor des Landesbetriebs für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt, Burkhard Henning, sagte dem MDR am Montagabend dazu, man gehe davon aus, dass die Alarmstufe 2 nur angekratzt werde. Das hänge auch mit der Steuerung der Wehranlagen zusammen. Es gelte "wachsam sein, aber keine dramatische Situation", so Henning.

Auch an der Schwarzen Elster könnte aus seiner Sicht auf Grund der Vorhersagen und den Informationen aus Brandenburg Alarmstufe 2 "gekratzt"werden. Der Fluss werde sehr stark durch die Tagebaue beeinflusst. Auch hier gelte es, wachsam zu sein. Es sei aber wie an der Ohre nichts Dramatisches.

Harz: Situation steuerbar

An der Bode habe das Rappbodetalsperrensystem viel Wasser zurückgehalten, erklärte der Experte mit Blick auf den Harz. Man könne die Situation gut aussteuern und werde auch mit den angekündigten Niederschlägen gut zurechtkommen. Die Hochwasserzentrale rechnet damit, dass maximal untere Alarmstufen erreicht werden. Bei den kleineren Flüssen hängt es laut Henning davon ab, wie stark jeweils lokal der Regen fällt. Äber von einer dramatischen Situation gehen wir auch hier nicht aus", so Henning.

8.8.2 Generated Text

Leichte Sprache: In Sachsen-Anhalt gibt es immer noch viel Hochwasser | MDR.DE

In dem Bundes-Land Sachsen-Anhalt gibt es immer noch viel Hochwasser. Das ist auch in den Bundes-Ländern: • Sachsen-Anhalt, • Sachsen-Anhalt • und Thüringen so. Dort gibt es zum Beispiel die Stadt: Aue-Bad Schlema. In der Nähe von Aue-Bad Schlema gibt es einen Fluss. Der Fluss heißt: Schwarzer Elster. In den letzten Tagen ist dort viel Hochwasser passiert. Deshalb gibt es in vielen Orten in Sachsen-Anhalt jetzt besonders viel Hochwasser. Zum Beispiel: • In der Stadt Wolmirstedt, • in der Stadt Löben • und in der Stadt Ilsenburg. In diesen Orten gibt es Alarmstufe 1. Burkhard Henning ist der Direktor vom Landes-Betrieb für Hochwasser-Schutz und Wasser-Wirtschaft in Sachsen-Anhalt. Er hat der Presse gesagt: Es ist sehr wichtig, dass es in den Orten schnell genug Hochwasser gibt. Deshalb ist es wichtig, dass es in den Orten schnell genug Hochwasser gibt.

8.9 Article Pair 9

8.9.1 Source Text

2. Bundesliga: Shpilevski wird neuer Cheftrainer von Erzgebirge Aue | MDR.DE

Ein langjähriger Nachwuchstrainer von RB Leipzig, der anschließend Kairat Almaty zum kasachischen Meistertitel führte, wird Dirk Schusters Nachfolger auf der Trainerbank von Erzgebirge Aue. Die Wahl des Zweitligisten fiel auf den erst 33-jährigen Aleksey Shpilevski.

Auf der Suche nach einem neuen Trainer ist Zweitligist Erzgebirge Aue in Kasachstan fündig geworden. Der 33-jährige Aleksey Shpilevski kommt von Kairat Almaty und erhält einen Dreijahresvertrag bis 2024, der für die 2. Bundesliga und Bundesliga gilt. Der gebürtige Weißrusse war im Alter von sechs Jahren nach Deutschland gekommen und spielte im Nachwuchs des VfB Stuttgart, ehe er mit nur 18 Jahren seine aktive Laufbahn verletzungsbedingt beenden musste.

Leonhardt: Ïdentifiziert sich voll mit unseren Werten"

Äliaksei ist trotz seiner jungen Jahre bereits sehr erfahren", wurde FCE-Präsident Helge Leonhardt in einer offiziellen Klubmitteilung zitiert und ergänzte: "Nach mehreren intensiven Gesprächen konnten wir feststellen, dass er sich mit den Werten unseres Vereins voll und ganz identifiziert. Ich freue mich, dass wir ihn zur Bewältigung der vor uns stehenden anspruchsvollen Aufgaben für unseren Club gewinnen konnten.SSeine Trainerlaufbahn trieb Shpilevski von 2013 bis 2018 im Nachwuchs von RB Leipzig voran. Die erste Aufgabe im Herrenbereich führte ihn anschließend für zwei Monate in sein Heimatland zu Dinamo Brest. Seit November 2018 verantwortete er den kasachischen Erstligisten Kairat Almaty, den er 2020 zum ersten Meistertitel seit 16 Jahren führte. Die offizielle Vorstellung des Fußballlehrers kündigte der Verein für übernächsten Dienstag 15. Juni an.

8.9.2 Generated Text

Leichte Sprache: Der Fußball-Verein Erzgebirge Aue will einen neuen Trainer | MDR.DE

Im Bundes-Land Kasachstan gibt es den Fußball-Verein Erzgebirge Aue. Er ist 33 Jahre alt. Und er spielt in der 2. Bundes-Liga. Der neue Trainer heißt: Aleksey Shpilevski. Er kommt aus dem Land Kasachstan. Und er hat einen Vertrag mit dem Fußball-Verein RB Leipzig. Er hat 3 Jahre einen Vertragals Trainer. Das bedeutet: Er spielt für 3 Jahre in einem Fußball-Verein. Und er bekommt bis zum Jahr 2024 einen Vertrag als Trainer. Das ist der Vertrag für die 2. Bundes-Liga und die Bundes-Liga. Alexey Shpilevski kommt aus dem Land Weißruss. Er ist 6 Jahre alt. Und er spielt in der 2. Bundes-Liga. Er hat den Vertrag mit dem Fußball-Verein RB Leipzig gemacht. Danach ist er zum Fußball-Verein Dynamo Brest gegangen. Das ist er seit dem Monat November im Jahr 2018 spielt. Er hat den Vertrag mit dem Fußball-Verein Erzgebirge Aue gemacht. Und er hat am Dienstag gesagt: Dass er den nächsten Trainer für RB Leipzig wird.

8.10 Article Pair 10

8.10.1 Source Text

Drei Schwerverletzte nach Unfall auf der A4 bei Dresden | MDR.DE

Bei einem Überholmanöver ist ein 45 Jahre alter Autofahrer auf der Autobahn 4 von der Fahrbahn abgekommen. Dabei wurden der Mann und seine beiden Mitfahrer schwer verletzt und kamen ins Krankenhaus. Der 45-Jährige war mit seinen 24 und 26 Jahre alten Insassen kurz vor Mitternacht auf der A4 von Eisenach in Richtung Dresden unterwegs, als er auf Höhe der Ausfahrt Dreieck Nossen bei einem Überholmanöver die Kontrolle über sein Auto verlor, teilte die Polizei am Freitag mit.

Demnach kam er nach rechts von der Fahrbahn ab und das Auto geriet in Brand. Während der Bergungs- und Rettungsmaßnahmen in der Nacht zum Freitag war die Autobahn zeitweise gesperrt.

8.10.2 Generated Text

Leichte Sprache: 45 Jahre alter Auto-Fahrer ist auf der Autobahn A4 von Dresden abgekommen | MDR.DE

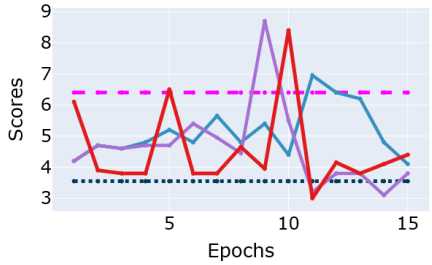
Im Bundes-Land Sachsen ist die Autobahn A4. Dort gab es in der letzten Nacht einen Unfall. So ist der Unfall passiert: Ein Auto-Fahrer ist auf der Autobahn A4 von der Stadt Eisenach in die Richtung von Dresden gefahren. Das ist eine Stadt im Bundes-Land Sachsen. Bei dem Unfall hat er sein Auto von der Fahrbahn abgekommen. Das bedeutet: Er ist von der Fahrbahn gefahren. Dabei hat er aber nicht mehr genug Kontrolle über sein Auto bekommen. Deshalb ist die Fahrbahn für einige Zeit gesperrt. Die Polizei hat der Presse gesagt: Bei dem Unfall wurden: • Der Auto-Fahrer • und mehrere andere Menschen verletzt. Sie wurden ins Kranken-Haus gebracht.

8.11 Automatic Metrics

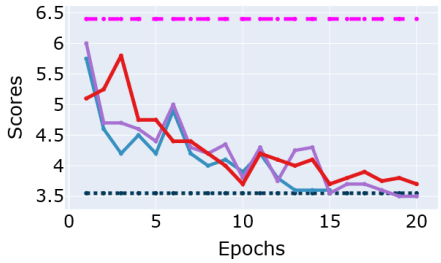
The figures given below show the progression of different metrics computed for each approach during fine-tuning for three runs. The best-performing run is always indicated in red. The reference scores for the three readability metrics are displayed using a dark blue dotted line and the input scores are displayed using a dashed magenta-colored line.

8.11.1 Median Wiener Sachtextformel Scores

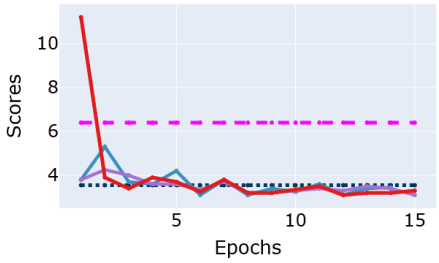
Figure 1: Median WSTF Scores During Fine-tuning



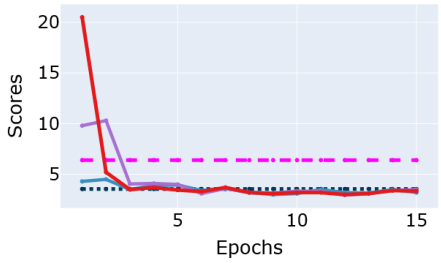
(a) mBART-sum Approach



(b) mBART-tr Approach



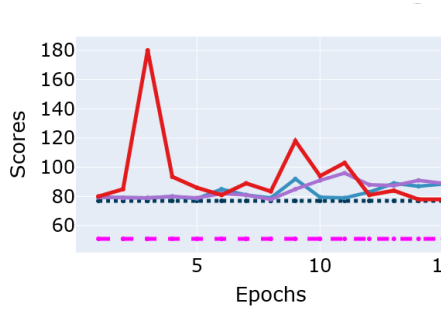
(c) mt5-sum Approach



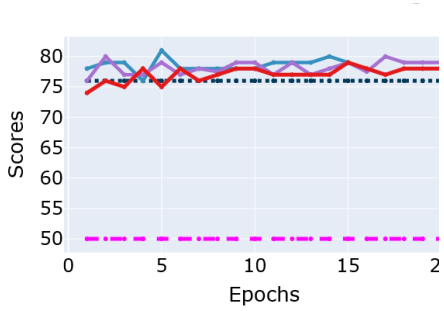
(d) mt5-tr Approach

8.11.2 Median Flesch-Reading-Ease Scores

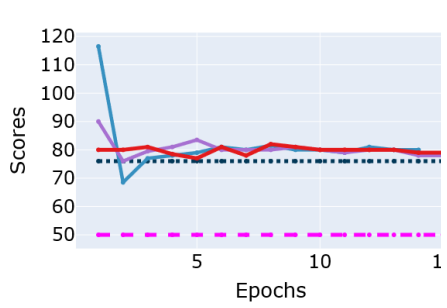
Figure 2: Median FRE Scores During Fine-tuning



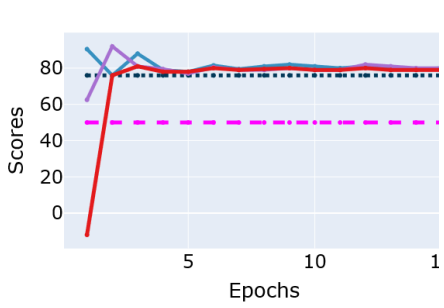
(a) mBART-sum Approach



(b) mBART-tr Approach



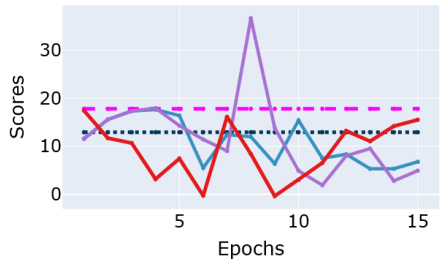
(c) mt5-sum Approach



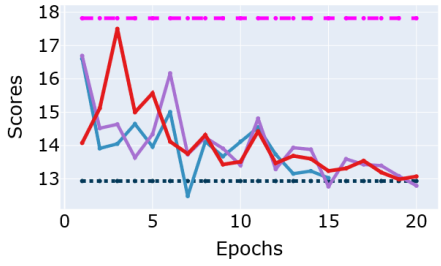
(d) mt5-tr Approach

8.11.3 Median Coleman-Liau Index Scores

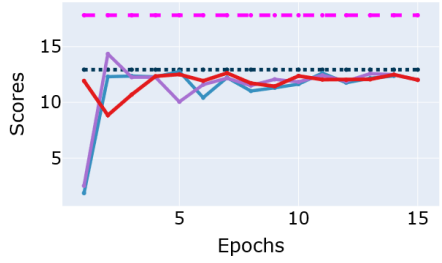
Figure 3: Median CLI Scores During Fine-tuning



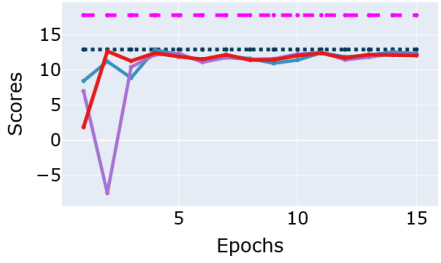
(a) mBART-sum Approach



(b) mBART-tr Approach



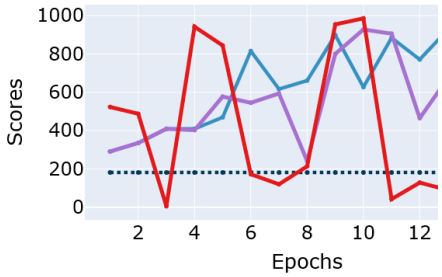
(c) mt5-sum Approach



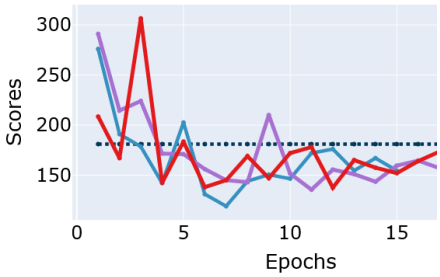
(d) mt5-tr Approach

8.11.4 Median Text Length Scores

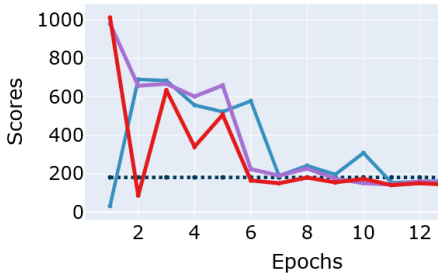
Figure 4: Median Text Length Scores During Fine-tuning



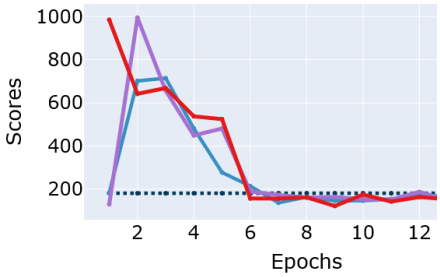
(a) mBART-sum Approach



(b) mBART-tr Approach



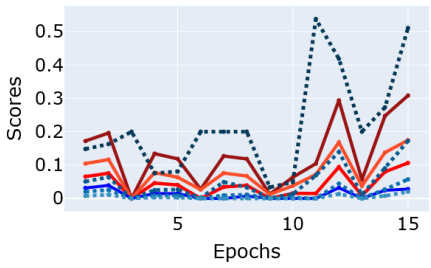
(c) mt5-sum Approach



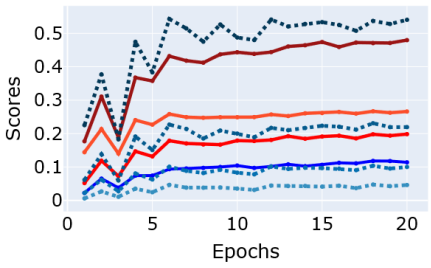
(d) mt5-tr Approach

8.11.5 BLEU and ROUGE Scores

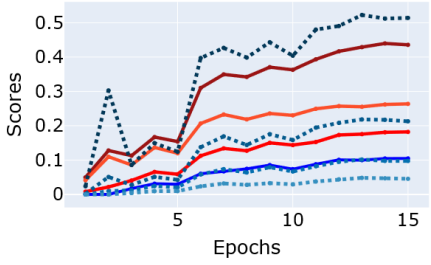
Figure 5: Median BLEU and ROUGE Scores During Fine-tuning



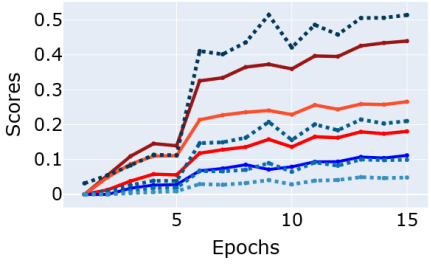
(a) mBART-sum Approach



(b) mBART-tr Approach



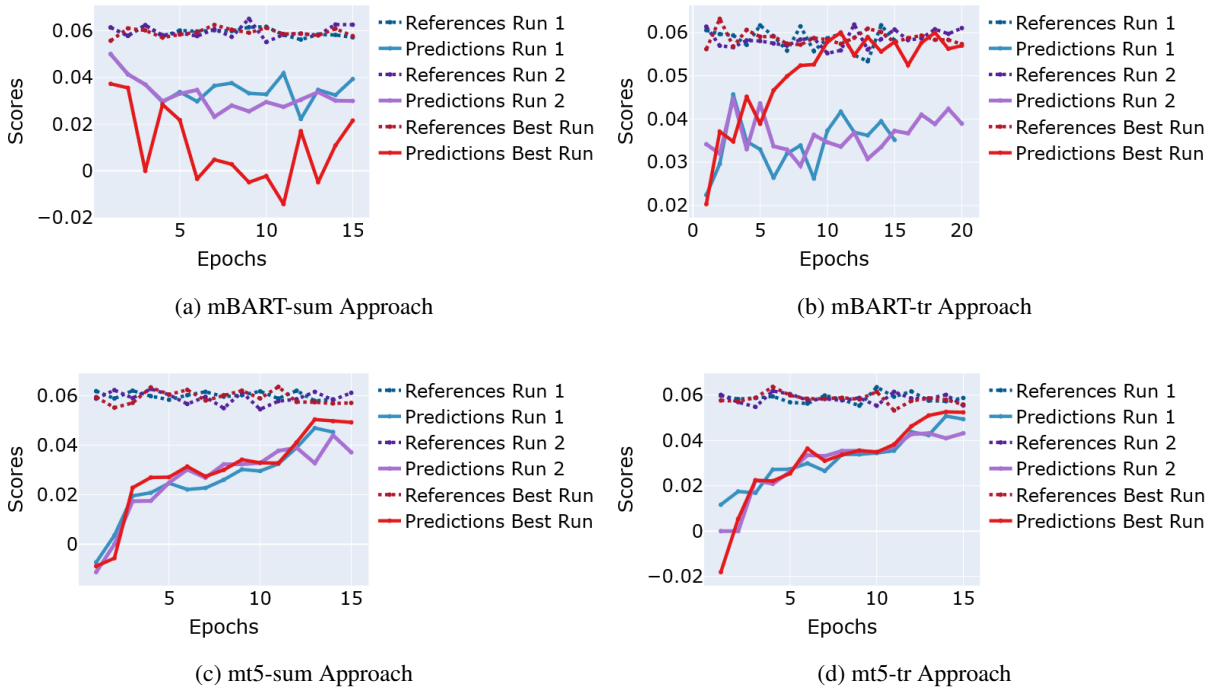
(c) mt5-sum Approach



(d) mt5-tr Approach

8.11.6 Median BLANC Scores

Figure 6: Median BLANC Scores During Fine-tuning



(a) mBART-sum Approach

(b) mBART-tr Approach

(c) mt5-sum Approach

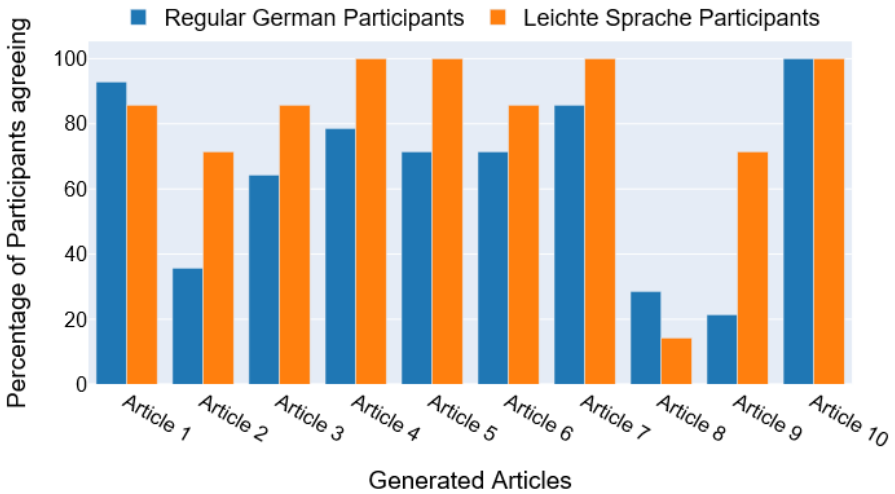
(d) mt5-tr Approach

8.12 Human Evaluation Responses

The following figures depict the responses of survey participants to six questions on the document-level simplification properties assessed. We use the term Leichte Sprache (LS) in the figures, which is the German term for German Easy-to-Read language or GE2R.

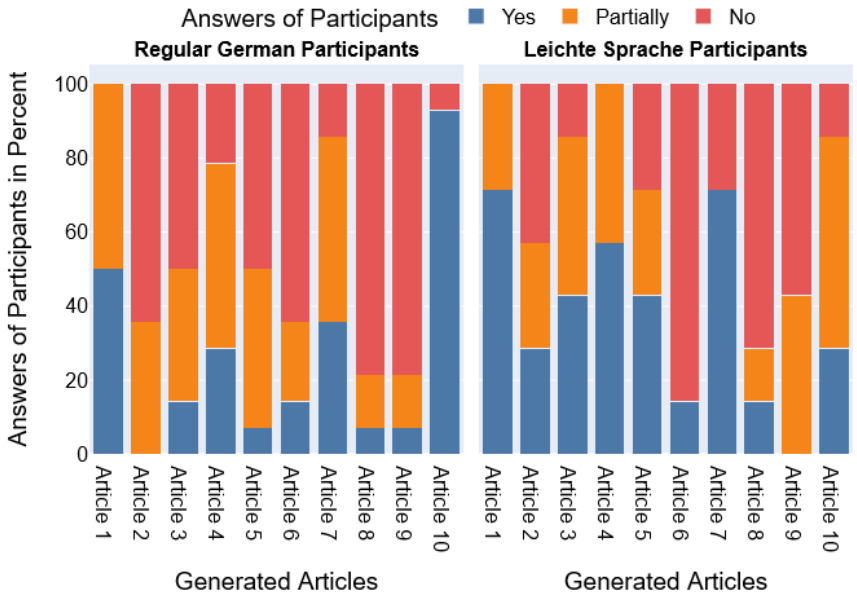
8.12.1 Intent Assessment

Figure 7: Responses to the Question: Does the generated article have the same intent as the source article?



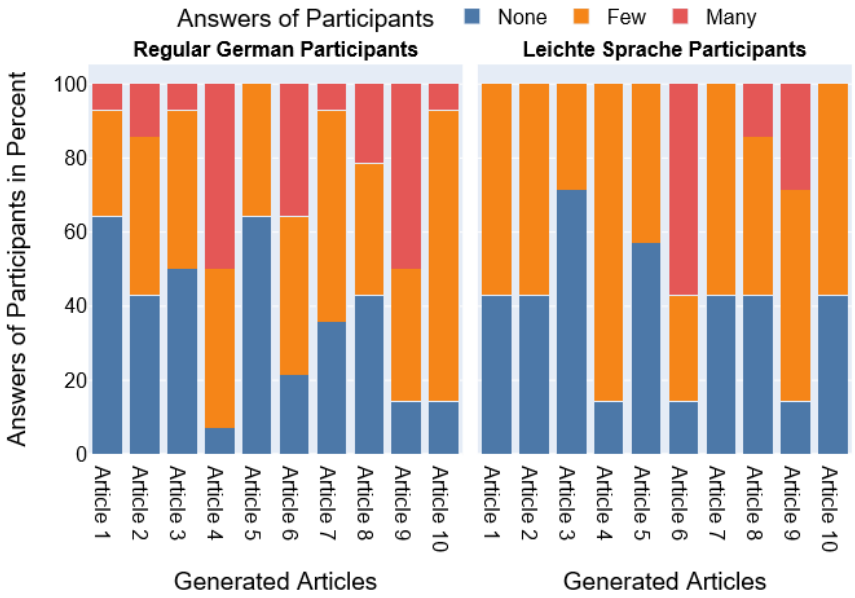
8.12.2 Content Assessment

Figure 8: Responses to the Question: Does the generated article represent the content of the source article well?



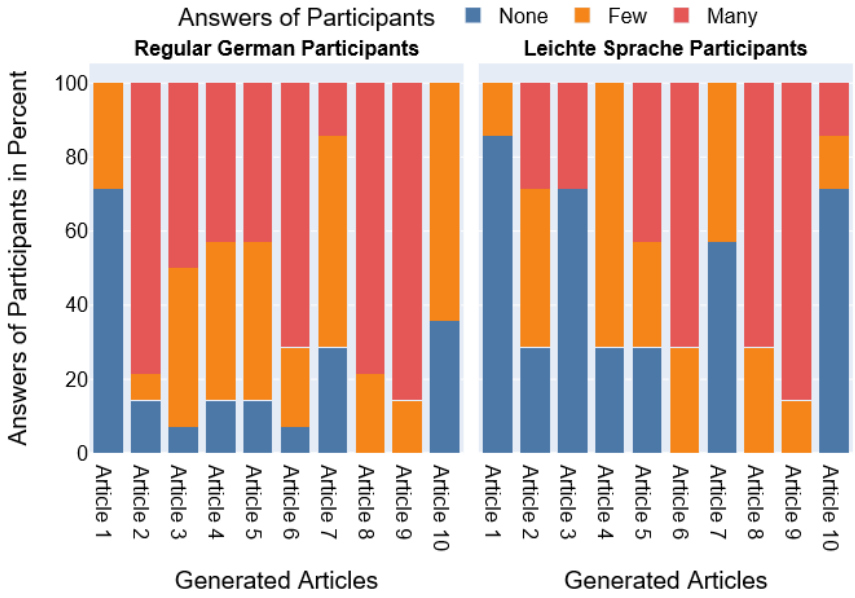
8.12.3 Grammar Assessment

Figure 9: Agreement of Participants with the Question: Does the generated article contain grammar errors? How many?



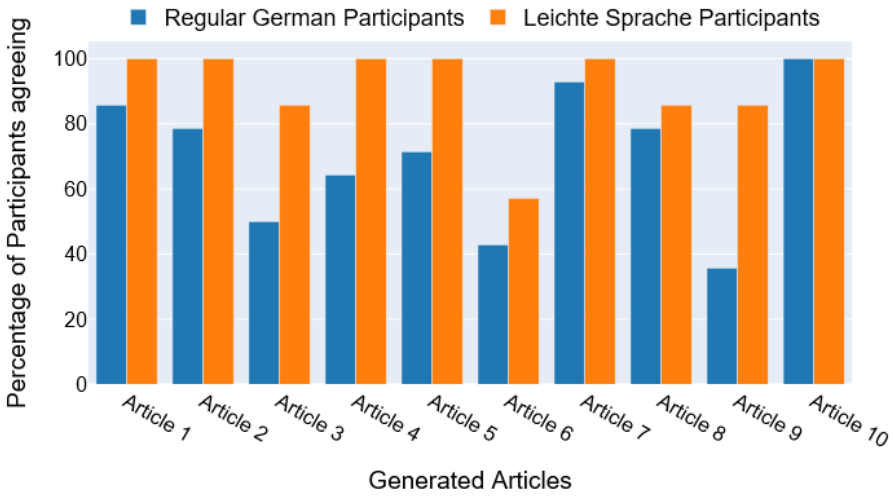
8.12.4 Logical Consistency Assessment

Figure 10: Responses to the Question: Does the generated article contain errors in its logical consistency? How many?



8.12.5 Reading Ease Assessment

Figure 11: Agreement of Participants with the Question: Is the generated article easy to read?



8.12.6 GE2R Language Assessment

Figure 12: Agreement of Participants with the Question: Is the generated article written in GE2R?