Automatic Generation of Comprehensible German Summaries for Domain-Specific Scientific Definitional Texts

Sophie Decher and Dr. Michael Dembach
Fraunhofer FKIE
Fraunhoferstraße 20
53343 Wachtberg, Germany

Abstract

This study aims to test a generative language model's capability to generate German comprehensible summaries of scientific definitional texts, while also gaining insight on which relevant linguistic features are part of the model's implicit knowledge. The open source multilingual model `mistral-small` was shown 100 texts from the EnArgus Wiki, a collection of German scientific definitional texts written by experts in the domain of energy research. The model was then prompted to generate two summaries for each of these texts. The first prompt specified that the texts should be in *einfacher Sprache*, or comprehensible German. The second prompt additionally listed a set of linguistic rules extracted from the DIN 8581-1 norm that the model was expected to comply with. The 200 automatically generated summaries were then assessed for sentence length, readability, semantic similarity to the original, and representation of the original text. A team of domain experts was asked to evaluate the summaries both qualitatively and quantitatively. In addition, two linguists assessed whether a subset of the generated summaries (n=100) adhered to the linguistic rules used in the generation prompt. The findings have implications for LLM use in science communication and NLG and provide insight on the difficulties of creating operationalizable guidelines for German comprehensible language.

1 Introduction

Public access to scientific knowledge and research is important. The high information density and low-frequency, specialized terminology of scientific texts (and especially the high percentage of compound words in German texts) can pose challenges for readers. These difficulties may stem from unfamiliarity with the topic at hand, reading level, or language barriers. In order to increase engagement with and understanding of current scientific research, science communication texts must be written in plain language.

"Translating" technical texts into plain language is not a simple task, however. Researchers, though deeply familiar with their own work, may still lack experience in science communication and knowledge of plain language guidelines—or simply the extra time—to create such texts and check that they meet the linguistic requirements. If a generative language model can be used for this application, science communication could benefit.

In the following, we take the EnArgus information portal [1] as the starting point for a case study on text generation of comprehensible scientific German texts. Funded by the German Federal Ministry of Economic Affairs and Energy, the EnArgus information system is designed to make German energy research funding more transparent and provide high-quality information on current research topics. Users—who range from federal ministries and project management organizations to scientific researchers and members of the general public—can search the databank of government funded projects or peruse the EnArgus Wiki, a collection of over 2000 definitional texts and the source of the data used in the experiments presented below. Written by domain experts, each of these wiki texts provides a definition for a topic relevant to energy research, e.g. *E-Bike*, *Dish-Sterling Kraftwerk* (Dish-Sterling power plant), or *Osmose* (osmosis). Though the authors strive for concision and clarity, the technical nature of many of the topics combined with the (compound-)noun-centricity of German means that the texts are likely not equally accessible for all readers.

Using a sample of these wiki texts, this study aims to investigate how well a multilingual generative language model can summarize German scientific definitional texts in comprehensible

[1] `https://enargus.de/`

Proceedings of the 21st Conference on Natural Language Processing (KONVENS), Volume 2: Workshops, pages 202–211, 2025

language. Supplementing each wiki text with such a summary would increase the accessibility of the EnArgus information system. In addition, the study aims to gain insight on whether the model relies on implicit linguistic knowledge during generation or needs explicit linguistic direction to be included during prompting. The following research questions are posed:

RQ$_1$: Is a multilingual generative language model capable of converting scientific definitional texts to comprehensible German summaries that comply with a series of pre-defined rules?

RQ$_2$: In how far does the quality of the generated texts improve when the linguistic rules are made explicit for the model?

The remainder of the paper is structured as follows. Section 2 provides an overview of comprehensible German and plain language, while Section 3 summarizes recent work on text generation and summarization in this area. In Section 4, the experimental setup is described. Section 5 presents the evaluation methods and corresponding results, which are discussed in further detail in Section 6. Concluding remarks and study implications can be found in Section 7.

2 Comprehensible German ("Einfache Sprache")

Comprehensible German texts may be written in either *leichte Sprache* or *einfache Sprache*. *Leichte Sprache* is a form of "barrier-free communication" with strict linguistic and typographic regulations, while *einfache Sprache* is defined less rigidly and places the focus on making texts understandable for the readers, whose needs may differ depending on the context and target audience in question (Bock and Pappert, 2023).

All references to "comprehensible German" in this paper refer to the term *einfache Sprache*, as the aim is to make scientific information more accessible for as wide a variety of readers as possible. The Deutsche Institut für Normung (DIN) has released a series of norms for comprehensible German. The *DIN ISO 24495-1 Einfache Sprache* norm is the German translation of the international norm for plain language and stresses the importance of addressing the needs of the particular target audience for a text (Deutsches Institut für Normung e. V., 2024). *DIN 8581-1 Anwendung für das Deutsche* lists specific linguistic rules and examples for writing texts in comprehensible German (Deutsches Institut für Normung e. V., 2025a). Most recently, the *DIN ISO 24495-3 Wissenschaftliche Sprache* norm, which includes some additional guidelines for scientific texts, was released as a draft (Deutsches Institut für Normung e. V., 2025b).

3 Text generation, summarization, and comprehensible language

The challenges that automatic generation of texts in comprehensible language poses are, as with any natural language generation task, two-fold: (1) model parameters and prompts must be optimized in order to obtain ideal output, and (2) model output must be appropriately evaluated. Unlike classification tasks, for which a gold standard dataset is easier to design, evaluating generative tasks is less straightforward. An additional level of complexity is added when domain-specific texts and texts in languages other than English are used as training data, as there are fewer resources available. A combination of qualitative and quantitative evaluation methods must be applied. A good example of methods that can be applied can be found in August et al. (2022), who analyzed function word usage, Flesch-Kincaid grade level, sentence length, and lexical choice in order to evaluate the complexity of the English generated scientific texts. Research on summarization of German texts, a closely related task for this study, is gradually increasing; while early efforts focused primarily on news summarization (Aumiller et al., 2023), more domains are now being considered, e.g. medicine (Ganzinger et al., 2025) or legal texts (Glaser et al., 2021).

4 Methodology

4.1 Data

The EnArgus Wiki [2] consists of over 2000 German scientific definitional texts written by experts in the domain of energy research. This study is based on a random sample of 100 wiki texts. In an initial preprocessing step, all text titles and metadata were removed so that only the main text body remained.

4.2 Model

All experiments were run on a 4-bit quantized version of the open source multilingual

[2] https://enargus.de/wiki/

`mistral-small` model [3]. The selection of this small model was deliberate, as not all scientific researchers or science communication experts have access to large servers. A smaller model is also less expensive to deploy and leaves a smaller energy footprint (Jurafsky and Martin, 2025). Testing a smaller, quantized model allows us to evaluate generative AI tools that are readily available without powerful computational resources. Temperature was set to zero to maximize the possibility of receiving deterministic responses and avoid hallucinations.

4.3 Prompt design and experimental setup

The model was prompted to create two summaries for each of the 100 original texts. The prompt for the first summary instructed the model to summarize the source text in 1-2 sentences in comprehensible German (see Appendix A for the full prompt). The text of the prompt for the second summary was identical to the first, with the addition of a list of nine linguistic rules selected from the DIN 8581-1 norm (Deutsches Institut für Normung e. V., 2025a; see Appendix B for the prompt). Figure 1 shows an example of an original text with its two corresponding summaries. This setup makes it possible to test whether the addition of explicit linguistic rules in the prompt results in a better summary, as compared to a prompt that only requests texts in comprehensible German. If the results from both prompts are similar, we can hypothesize that the rules are implicitly already part of the model's knowledge of comprehensible language. The resulting 200 automatically generated summaries were assessed automatically for sentence length (Rule 1 of the linguistic rules included in the prompt), readability score, and semantic similarity to the original. They were then rated and assessed by a group of domain experts. A manual linguistic evaluation was also done on a subset of the generated summaries (50 generated with linguistic rules and 50 without). For each of these 100 texts, compliance with Rules 2-9 was assessed. The evaluations are presented in Section 5.

4.4 Target audience and communicative goal for the generated texts

The aim of these experiments is to generate a concise summary of each source text in comprehensible German that gives a short overview of the main topic. The summaries should remain faithful to the corresponding source texts and be easily readable for members of the general public and laypeople. Domain-specific technical terms must be clearly explained.

[3] https://ollama.com/library/mistral-small:22b

5 Evaluation

5.1 Adherence to linguistic rules

In order to evaluate the model's understanding of Rules 2-9, the first 100 generated summaries (50 generated with the rules explicitly stated in the prompt, and 50 generated without the rules) were manually rated by two linguists.

Rule	Compliant texts generated with rules	Compliant texts generated without rules
2. Sentence structure	4	22
3. Subordination I	48	50
4. Subordination II	49	50
5. Genitive attributes	48	50
6. Compound words I	48	48
7. Compound words II	48	49
8. Nominalizations	0	16
9. Abbreviations	48	48

Table 1: Linguistic rule evaluation for rules 2–9: Compliancy with linguistic rules in texts generated with linguistic rules stated explicitly in the prompt (n=50) and texts generated without explicit statement of the rules in the prompt (n=50).

Length (words)	Original texts	Texts generated with rules	Texts generated without rules
Mean text length	192.81	120.84	56.80
Min text length	22.00	15.00	18.00
Max text length	516.00	325.00	123.00
Mean sentence length	13.80	11.11	15.06
Min sentence length	3.00	3.00	3.00
Max sentence length	50.00	38.00	34.00

Table 2: Text and sentence length in words in original texts (n=100) and texts generated both with rules stated in the prompt (n=100) and without (n=100). The sentence splitter identified a few sentences that were shorter than three words. These were excluded from this analysis, as they were subheadings rather than full sentences.

The evaluated rules were selected from a list of 114 linguistic rules relevant to the task at hand compiled from the guidelines in the DIN 8581-1 norm (99 rules) and the most recent draft of the DIN ISO 24495-3 norm (5 rules). The DIN 8581-1 norm proved most useful for this task, as it contained formalized, specific rules at multiple levels

Summaries for the text *Stoffaufbereitung*

Original text

Die Stoffaufbereitung ist ein wichtiger Prozessschritt der Papierherstellung. Da sowohl die Produktion auf der Papiermaschine als auch das fertige Produkt genaue Anforderungen an die Eigenschaften des eingesetzten Stoffgemischs stellen, ist es Ziel und Zweck der Stoffaufbereitung, den eingesetzten Faserstoff aufzubereiten und einen konstanten Stoffstrom für die Papiermaschine zur Verfügung zu stellen. Dies ist insbesondere dann anspruchsvoll, wenn die Eigenschaften der eingesetzten Halbstoffe nicht zeitlich konstant sind. Abhängig von den Anforderungen an das Papier werden unterschiedliche Prozessschritte in der Stoffaufbereitung eingesetzt. Zunächst wird der Faserstoff in einem Pulper in Wasser aufgelöst. Um die für die jeweilige Papiersorte benötigten Eigenschaften des Faserstoffs zu garantieren, wird er häufig in der Papierfabrik in einem Refiner nachgemahlen. Je nach Fasermaterial und Papiersorte kann dieser Produktionsschritt sehr energieintensiv sein. Weitere Prozessschritte der Stoffaufbereitung können die Sortierung, Reinigung, erneute Nachmahlung und Dispergierung, also das Einarbeiten von Füllstoffen in die Faser-Wasser Suspension sein.

Stock preparation is an important process step in papermaking. Since both production on the paper machine and the finished product place precise demands on the properties of the material mixture used, the aim and purpose of stock preparation is to prepare the fiber used and provide a constant material flow for the paper machine. This is particularly challenging when the properties of the pulp used are not constant over time. Depending on the requirements of the paper, different process steps are used in stock preparation. First, the fiber is dissolved in water in a pulper. To guarantee the fiber properties required for the respective paper grade, it is often refined in a refiner at the paper mill. Depending on the fiber material and paper grade, this production step can be very energy-intensive. Further process steps in stock preparation can include screening, cleaning, further refining, and dispersion, i.e. the incorporation of fillers into the fiber-water suspension.

Summary generated without rules

Die Stoffaufbereitung ist ein wichtiger Schritt bei der Papierherstellung. Dabei wird der Faserstoff in Wasser aufgelöst und gegebenenfalls gemahlen, um die gewünschten Eigenschaften des Papiers zu erreichen. Dieser Prozess kann energieintensiv sein und beinhaltet oft auch Sortierung, Reinigung und das Einarbeiten von Füllstoffen.

Stock preparation is an important step in papermaking. This involves dissolving the fiber in water and, if necessary, refining it to achieve the desired paper properties. This process can be energy-intensive and often includes screening, cleaning, and the incorporation of fillers.

Summary generated with rules

Die Stoffaufbereitung ist ein wichtiger Schritt bei der Papierherstellung. Das Ziel ist es, den eingesetzten Faserstoff aufzubereiten und einen konstanten Stoffstrom für die Papiermaschine bereitzustellen. Der Faserstoff wird in einem Pulper in Wasser gelöst. Er kann auch in einem Refiner nachgemahlen werden, um die benötigten Eigenschaften zu erreichen. Dieser Schritt kann sehr energieintensiv sein, je nach verwendetem Fasermaterial und Papiersorte. Weitere Schritte sind Sortieren, Reinigen, Nachmahlen und Dispergieren, also das Einarbeiten von Füllstoffen in die Faser-Wasser-Mischung

Stock preparation is a key step in papermaking. The goal is to process the fiber input and provide a consistent flow of material for the paper machine. The fiber is dissolved in water in a pulper. It can also be reground in a refiner to achieve the required properties. This step can be very energy-intensive, depending on the fiber material used and the paper grade. Further steps include screening, cleaning, reground, and dispersing, i.e., incorporating fillers into the fiber-water mixture.

Figure 1: Summaries of the text *Stoffaufbereitung* generated with and without linguistic rules in the prompt

of linguistic analysis. Surprisingly, the DIN ISO 24495-3 norm, though specified to be applicable for scientific writing in comprehensible German, did not contain many additional formalized rules, instead repeating many of the more general guidelines in DIN ISO 24495-1 with added emphasis on the importance of defining the target audience of a text. The five rules taken from DIN ISO 24495-3 touched on comparisons, noun phrases, clear sentences, replacement of technical terms with simpler synonyms, and avoidance of hedges, i.e. vague or ambiguous terms such as *ähnlich* (similar), *nahe* (near), *einen Trend* (a trend), *näherungsweise* (approximately). No rules were extracted from the DIN ISO 24495-1, as the guidelines in this norm were originally designed for English and were covered again in more detail in the DIN 8581-1 norm. From the master list of possible rules to evaluate, the nine rules listed below, all taken from DIN 8581-1, were selected for the linguistic evaluation because each presents a clear formalization of a linguistic phenomenon on the word or sentence level that can be evaluated on a binary (yes/no) basis (Rules 2-9) or automatically (Rule 1). They make specific features of comprehensible language explicit. The sentence and word levels are important for this task, as the goal is to generate very short summarizations of longer texts, requiring optimized lexical choice and tight control of sentence length and structure. The word level is particularly relevant, as the original texts contain difficult, infrequently-occurring scientific terminology. Other rules from the DIN norms such as "The text should be grammatically correct" are assumed to be part of the implicit knowledge of the model and are therefore not included in the evaluation for this study.

Table 1 shows the results for Rules 2-9, which were manually evaluated. Rule 1 was evaluated automatically. The findings for all rules as well as their general relevance for the task are discussed below.

1. **Sentence length: Die Satzlänge soll variiert werden und nie 15 Wörter überschreiten.** *Sentence lengths should vary and never exceed 15 words.* Neither text group adhered completely to this rule. Texts generated without rules tended to consist of fewer sentences (M = 3.87 sentences) that were longer (M = 15.20 words per sentence; cf. Table 2). 88 of these texts contained at least one sentence that was longer than 15 words. Conversely, texts generated with rules were made up of more sentences (M = 10.91 sentences) but had fewer average words per sentence (M = 11.31 words). 65 of these texts contained at least one sentence longer than 15 words. Prompts for both sets of texts set a length limit of 1-2 sentences for the generated summaries, in order to obtain concise, brief summaries. The model was not able to comply with this part of the prompt, though performance was better for the set of texts generated without rules. Length limits in future generative prompts should be determined by word count rather than sentence count.

   The importance of this rule for comprehensible German texts must also be discussed. Wolfer et al. (2015) confirm that simplifying German texts by shortening the sentences often results in an increase in referential expressions such as pronouns and demonstratives that serve to maintain coherence across sentences. These expressions, in turn, require more parsing on the part of the reader and may increase reading difficulty. Finding a balance between sentence length and referential mechanisms may be more beneficial for comprehensible German texts.

2. **Sentence structure: Alle Sätze sollten der kanonischen Satzstruktur in Form von Subjekt–Verb–Objekt (SVO) entsprechen.** *All sentences should conform to the canonical sentence structure in the form of subject–verb–object (SVO).* As Table 1 shows, only four texts generated with rules complied with this rule, whereas 22 texts generated without rules were compliant. In German, this rule can only be applied to main clauses (Wöllstein, 2010; Eisenberg, 2020b). Due to this word order limitation, it must be considered whether a rule requiring SO order, regardless of the placement of the verb, would be more fitting for comprehensible German texts. When a German main clause starts with an adverbial, the subject is pushed to the position behind the verb (Wöllstein, 2010). This was the most common construction in the generated summaries that failed to comply with this rule. The change in word order often serves to make a reference to the preceding sentence in the text. If changes in word order

maintain the cohesion of a text, it might be argued that they enhance rather than detract from the comprehensibility of that text.

3. **Subordination I: Ein Satz soll höchstens einen Nebensatz haben.**
   Sentences should contain no more than one subordinate clause. Sentences with more than one subordinate clause were rare. Only two of the evaluated texts, both generated with the prompt that explicitly listed the rule, violated this rule. As all evaluated texts generated without the rules were compliant, it appears that the model does not need to be explicitly prompted to limit subordination.

4. **Subordination II: Ein Nebensatz kann vor oder nach dem Hauptsatz stehen.**
   A subordinate clause may precede or follow the main clause. Only one of the evaluated texts violated this rule. It, too, was generated with the prompt that included the rule explicitly, suggesting that prompting with this rule is counterproductive.

5. **Genitive attributes: Genitivattribute sollen nachgestellt werden, es sei denn, es handelt sich um Eigennamen.**
   Genitive attributes should be placed after the noun, except in the case of proper nouns. With the exception of two texts generated with rules, this rule was successfully followed (see Table 1). These two violations involved relative pronoun constructions, which are not necesesarily explicitly covered by this rule.

   Of note is the fact that texts generated with rules contained more genitive constructions than those generated without the rules. Adding a rule about genitive constructions in the generation prompt seems to have primed the model to use the genitive case more often. As the positioning of genitive attributes in German sentences already usually defaults to after the noun, the model is likely implicitly aware of this rule, rendering it an unnecessary part of the prompt.

6. **Compound words I: Überlange Komposita (länger als 3 Lexeme) sollten entweder mit Bindestrich geschrieben oder aufgelöst werden.**
   Overly long compound words (longer than 3 lexemes) should either be hyphenated or split up into separate words. In the summaries generated without rules in the prompt, two texts contained the overly long compound words *Pumpspeicherkraftwerk* and *Drehstromübertragungssystem*, respectively. The terms *Hochwasserentlastungsanlage* and *Dampfstrahlkälteanlagen* were also included in two different texts that were generated with the rules in the prompt. These technical scientific terms seem to be especially challenging for the model to simplify. This is a task that must be addressed in order to ensure that German scientific texts are understandable for all readers.

7. **Compound words II: Wenn Komposita mit Bindestrich geschrieben werden, dann soll dieser an der Hauptfuge gesetzt werden.**
   If compound words are hyphenated, this should occur at the main joint (primary compound boundary). When overly long compound words occurred, they were never hyphenated incorrectly. Violations of this rule were due to the fact that no hyphenation was employed at all. The importance of this rule is secondary to that of the other rules, at least in the context of the data evaluated here.

8. **Nominalizations: Substantivierungen sollen vermieden werden. Stattdessen sollten Verbalkonstruktionen verwendet werden.**
   Nominalizations should be avoided. Verbal constructions should be used in their place.

   It is not clear which types of nominalizations this rule is meant to include. If, as stated, no nominalizations of any kind are to be permitted, almost all 200 texts violate the rule, as can be seen in Table 1. This would mean, however, that very common words like *Herstellung* would not be allowed in comprehensible German texts. An alternative would be to disallow only nominalizations that have been created via the process of conversion, i.e. transformed into a new word class without any change in form, such as in the case of *trinken* vs. *das Trinken* (Eisenberg, 2020a). Very few of the 200 generated summaries include nominalizations of this type. From a reading comprehension standpoint, it also makes sense to avoid word forms whose word class might be ambiguous for the reader when writing or

generating comprehensible texts.

9. **Abbreviations: Unbekannte oder fachsprachliche Abkürzungen müssen im Text eingefuhrt beziehungsweise erklärt werden.**
   Uncommon or technical abbreviations must be introduced or explained in the text. This rule could not be evaluated effectively, as instances of abbreviations in the generated summaries were very rare overall.

5.2 Similarity to original text

The similarity of the 200 generated summaries to their 100 original counterparts was assessed in two ways. First, all texts were embedded using the multilingual `paraphrase-multilingual-mpnet-base-v2`[4] sentence transformers model (Reimers and Gurevych, 2019). The cosine similarity between each generated summary and the corresponding source text was then calculated. As can be seen in Figure 3, the majority of the summaries generated without rules in the prompt were less similar to the source text as compared to the summaries generated with rules in the prompt. For the second assessment, a group of 7 domain experts were asked to rate the 200 generated summaries in a blind binary comparison task. Each text was evaluated once. Evaluators were presented with an original text and the two corresponding generated summaries in randomized order and asked "Which of these two texts is a better summary of the original?" The summaries generated without the rules in the prompt were selected more often (71%) than those generated with the rules in the prompt (29%). When asked to comment on their choices, raters criticized the apparent hallucination and inaccuracies in some of the generated summaries (see Figure 2), as well as the unnecessary retention of some complicated technical terms such as *Methyl-Tertiär-Butyl-Ether*. They often noted that the summaries generated without rules were slightly too short, missing key information, or lacking in clarity and flow, while the summaries generated with rules were often too long and in depth to be considered summaries.

[4] `https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2`

Excerpt from a summary of *Dezentrale Trinkwassererwärmung*

...Allerdings sind die Kosten für dezentrale Systeme oft höher als für zentrale. Ein weiterer Vorteil ist, dass die Heizungsanlage effizienter betrieben werden kann, besonders im Sommer....

...However, the costs for decentralized systems are often higher than for centralized ones. Another advantage is that the heating system can be operated more efficiently, especially in summer....

Figure 2: Incorrect information in an excerpt from a summary of the EnArgus Wiki text *Dezentrale Trinkwassererwärmung* generated with linguistic rules in the prompt. Higher costs are not advantageous, as this summary suggests.

5.3 Readability score

For an additional assessment method, the readability scores of the original texts and the generated summaries were calculated. As the Flesch-Kincaid grade level used by August et al. (2022) is not as accurate for German as it is for English, we chose the Wiener Sachtextformel IV (Wild and Pissarek, 2021) as an alternative, as shown in Equation 1:

$$\text{WSTF} = 0.2656 \times \left(\frac{\text{\# words}}{\text{\# sentences}}\right) + 0.2744 \times \left(\frac{\text{\# words longer than 3 syllables}}{\text{\# words}}\right) \times 100 - 1.693 \quad (1)$$

Syllables were counted using the `pyphen`[5] Python package, while sentences and words were counted with the `spacy`[6] package using the `de_core_news_sm` pipeline. Figure 4 shows that the readability scores for the summaries generated without rules were overall higher than those of the summaries generated with rules. The original source texts still had the highest readability score, however.

6 Discussion

In light of the above analyses, RQ_1 can partially be confirmed. The manual evaluation of generated summaries according to the linguistic rules showed that most texts complied with the rules for subordination, genitive attributes, and abbreviations. They also complied with the rules for compound words, with some exceptions. Clearly, transforming scientific texts into comprehensible German requires a

[5] `https://pyphen.org/`
[6] `https://spacy.io/`

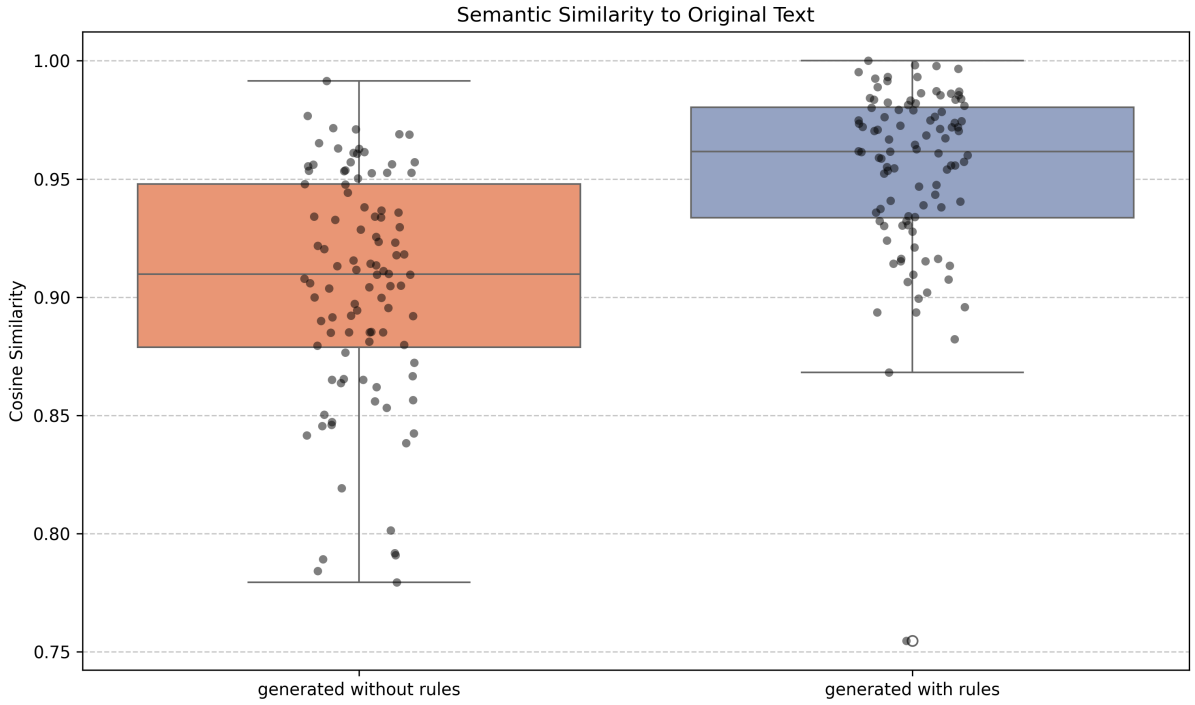

Figure 3: Similarity of generated summaries (both with and without explicit linguistic rules in the generation prompt) to their respective original texts

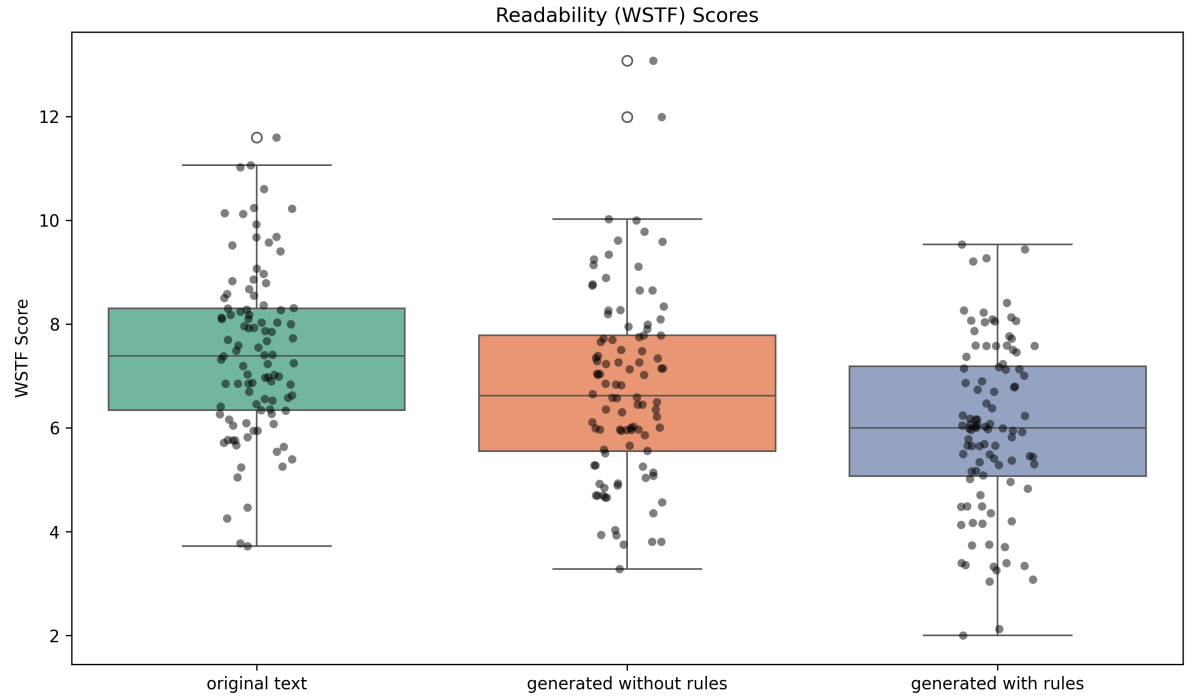

Figure 4: Readability scores of original texts and corresponding summaries generated both with and without the inclusion of linguistic rules in the generation prompt. Scores computed using the Wiener Sachtextformel IV.

focus on the technical compound words typical for German. For all of these rules, there was often no difference between the summaries generated with and without the rules in the prompt.

However, three linguistic rules were not successfully followed. In the case of nominalization (Rule 8) and sentence structure (Rule 2), this is due to the strict formulation of the rules in the DIN 8581-1 norm. It may be necessary to further refine such guidelines for authors, especially when it comes to creating comprehensible versions of scientific technical texts. In the case of sentence length (Rule 1), texts generated both with and without explicit statement of the linguistic rules contained sentences longer than 15 words. We hypothesize that, when prompted with linguistic rules, the model generated texts with a larger overall word count and a higher number of sentences, as it was trying to comply with all rules. This allowed the sentences to be shorter on average. In contrast, texts generated without explicit rules had a smaller word count and were made up of fewer sentences, which resulted in an increase in sentence length. The additional length requirement of 1-2 sentences in both prompts added further constraints–in future studies, it might be better to specify length requirements with a word count rather than a sentence count.

The explicit inclusion of linguistic rules during prompting did not improve the quality of the generated texts (RQ_2). In terms of nominalizations and sentence structure, quality worsened when the rules were explicitly stated. The inclusion of the remaining linguistic rules in the prompt during generation seems to have had little to no effect in most cases. It may be that the rules pertaining to subordination, genitive attributes, compound words, and lexical choice are either already part of the model's implicit linguistic knowledge of what "einfache Sprache" entails or unimportant for generation. The amount of genitive constructions overall also increased when the corresponding rule was included in the generation prompt. The inclusion of rules seems to have influenced the model in unexpected ways.

Readability scores and domain expert ratings were better for texts generated without the rules, though this group of summaries was less semantically similar to the corresponding source texts than the group of summaries generated with rules. It is clearly important to strike a balance between faithfulness to the original source text and comprehensibility for the reader, particularly for this domain of texts.

It seems that the model had difficulties generating texts according to the prompt that included the linguistic rules. It is not clear whether the prompt was too long or too complicated for the model to understand. Previous research on prompt engineering stresses the importance of structure, brevity, coherence, and clarity in prompts (cf. the CLEAR Framework, Lo, 2023). It is possible that the model, in trying to fulfill some of the rules listed, was then prevented from fulfilling other rules listed.

7 Conclusion

Both human ratings and automated evaluations of the generated summaries indicate that the quality of the summaries as-is is inadequate for immediate public use. Generating comprehensible versions of German scientific definitional texts exacerbates the challenges of general summarization. More experiments and model comparisons are needed to continue this line of research.

Limitations

This study has a number of limitations that affect the generalizability of the results. Due to limited resources and a small pool of human raters, only a small subcorpus of data could be used to prompt the model and generate summaries. Similarly, it was only possible to evaluate a small number of linguistic rules from the DIN norms. Follow-up studies are needed to examine a larger quantity of data. For improved summary quality, larger generative models without quantization should also be tested and compared.

Acknowledgments

This study is funded by the German Federal Ministry for Economic Affairs and Energy, as part of the EnArgus 3.0 project (Förderkennzeichen 03EI1062B).

References

Tal August, Katharina Reinecke, and Noah A. Smith. 2022. Generating scientific definitions with controllable complexity. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers*. Association for Computational Linguistics.

Dennis Aumiller, Jing Fan, and Michael Gertz. 2023. On the state of German (abstractive) text summarization. *arXiv:2301.07095*.

Bettina M. Bock and Sandra Pappert. 2023. *Leichte Sprache, Einfache Sprache, verständliche Sprache*. Narr Francke Attempto Verlag, Tübingen, Germany.

Deutsches Institut für Normung e. V. 2024. DIN ISO 24495-1:2024-03 – Einfache Sprache – Teil 1: Grundsätze und Leitlinien.

Deutsches Institut für Normung e. V. 2025a. DIN 8581-1 Einfache Sprache - Anwendung für das Deutsche - Teil 1.

Deutsches Institut für Normung e. V. 2025b. DIN ISO 24495-3:2025-05 – Entwurf – Einfache Sprache – Teil 3: Wissenschaftliches Schreiben.

Peter Eisenberg. 2020a. *Grundriss der deutschen Grammatik. Das Wort*, 5th edition. J. B. Metzler, Berlin, Germany.

Peter Eisenberg. 2020b. *Grundriss der deutschen Grammatik. Der Satz*, 5th edition. J. B. Metzler, Berlin, Germany.

Matthias Ganzinger, Nicola Kunz, Pascal Fuchs, Cornelia K. Lyu, Martin Loos, Martin Dugas, and Thomas M Pausch. 2025. Automated generation of discharge summaries: Leveraging large language models with clinical data. *Scientific Reports*, 15(16466).

Ingo Glaser, Sebastian Moser, and Florian Matthes. 2021. Summarization of German court rulings. In *Proceedings of the Natural Legal Language Processing Workshop 2021*, pages 180–189, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd edition. Online manuscript released January 12, 2025.

Leo S. Lo. 2023. The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Johannes Wild and Markus Pissarek. 2021. Ratte 2: Regensburger Analysetool für Texte - Dokumentation.

Sascha Wolfer, Sandra Hansen-Morath, and Lars Konieczny. 2015. Are shorter sentences always simpler? Discourse level processing consequences of reformulating jurisdictional texts. In Karin Maksymski, Silke Gutermuth, and Silvia Hansen-Schirra, editors, *Translation and Comprehensibility*, pages 263–287. Frank & Timme GmbH, Berlin, Germany.

Angelika Wöllstein. 2010. *Topologisches Satzmodell*. Universitätsverlag Winter Heidelberg, Heidelberg, Germany.

Appendices

A Prompt without linguistic rules

```
Du bist ein Experte der Wissenschaftskommunikation und seit Jahren im Bereich der Energieforschung tätig. Fasse diesen wissenschaftlichen Lexikonartikel in einfacher Sprache in 1-2 Sätzen zusammen. Das erste Substantiv im ersten Satz soll nicht geändert werden. Hier ist der Text: \n\n {input_text}
```

B Prompt with linguistic rules

```
Du bist ein Experte der Wissenschaftskommunikation und seit Jahren im Bereich der Energieforschung tätig. Fasse diesen wissenschaftlichen Lexikonartikel in einfacher Sprache in 1-2 Sätzen anhand der folgenden Regeln zusammen:

1. Die Satzlänge soll variiert werden und nie 15 Wörter überschreiten.
2. Alle Sätze sollten der kanonischen Satzstruktur in Form von Subjekt-Verb-Objekt (SVO) entsprechen.
3. Ein Satz soll höchstens einen Nebensatz haben.
4. Ein Nebensatz kann vor oder nach dem Hauptsatz stehen.
5. Genitivattribute sollen nachgestellt werden, es sei denn, es handelt sich um Eigennamen.
6. Überlange Komposita (länger als 3 Lexeme) sollten entweder mit Bindestrich geschrieben oder aufgelost werden.
7. Wenn Komposita mit Bindestrich geschrieben werden, dann soll dieser an der Hauptfuge gesetzt werden.
8. Substantivierungen sollen vermieden werden. Stattdessen sollten Verbalkonstruktionen verwendet werden.
9. Unbekannte oder fachsprachliche Abkurzungen mussen im Text eingefuhrt beziehungsweise erklärt werden.

Das erste Substantiv im ersten Satz soll nicht geändert werden. Hier ist der Text:\n\n {input_text}
```