Ethics and Information Technology (2024) 26:50
https://doi.org/10.1007/s10676-024-09792-4

ORIGINAL PAPER

Easy-read and large language models: on the ethical dimensions of LLM-based text simplification

Nils Freyer[1,2] · **Hendrik Kempt**[3] · **Lars Klöser**[2]

Accepted: 27 July 2024 / Published online: 4 August 2024
© The Author(s) 2024

Abstract

The production of easy-read and plain language is a challenging task, requiring well-educated experts to write context-dependent simplifications of texts. Therefore, the domain of easy-read and plain language is currently restricted to the bare minimum of necessary information. Thus, even though there is a tendency to broaden the domain of easy-read and plain language, the inaccessibility of a significant amount of textual information excludes the target audience from partaking or entertainment and restricts their ability to live life autonomously. Large language models can solve a vast variety of natural language tasks, including the simplification of standard language texts to easy-read or plain language. Moreover, with the rise of generative models like GPT, easy-read and plain language may be applicable to all kinds of natural language texts, making formerly inaccessible information accessible to marginalized groups like, a.o., non-native speakers, and people with mental disabilities. In this paper, we argue for the feasibility of text simplification and generation in that context, outline the ethical dimensions, and discuss the implications for researchers in the field of ethics and computer science.

Keywords Large language models · Easy read · AI ethics · Natural language processing · Accessibility

Introduction

Text sources play a crucial role in distributing a wide range of information and therefore, accessing language is an important support to individual autonomy and justice. Not only does the ability to access language in itself increase one's autonomy but the extent of procedural knowledge accessible via spoken or written language, especially online.

Especially marginalized persons such as non-native speakers and people with learning- or mental disabilities are affected by the exclusion by language (Cheung, 2017; Jones & Williams, 2017). For instance, from a perspective of distributive justice, inaccessible language threatens the principle of equality of opportunity (Rawls, 1971), as language is an important means to education and procedural knowledge. Similarly, from a perspective of democratic egalitarianism and relational justice, language that is not accessible to all affects the virtue of mutual moral equality (Anderson, 1999). Text simplification methods such as easy-read and plain language aim to make written or spoken language easier to understand for these groups. However, creating simplified texts requires well-educated and sensibly trained experts who can understand and empathize with different marginalized groups and levels of language comprehension (Rink, 2023). As a result, while many public institutions want to generate and translate texts in this format, the amount of accessible information is mostly limited to administrative details, making it difficult for these groups to access other textual information autonomously.

The emergence of services such as ChatGPT has brought generative AI (GenAI) and more specifically large language models (LLMs) into the spotlight of ethical debates. These

✉ Nils Freyer
nfreyer@ukaachen.de

Hendrik Kempt
hendrik.kempt@humtec.rwth-aachen.de

Lars Klöser
kloeser@fh-aachen.de

[1] Department of Medical Informatics, University Hospital, RWTH Aachen University, Aachen, Germany

[2] FH Aachen - University of Applied Sciences, Aachen, Germany

[3] Applied Ethics, RWTH Aachen University, Aachen, Germany

powerful models have the potential to automate a wide range of natural language tasks with relative ease.[1]

By leveraging LLMs, generating or translating simplified language texts could become much less challenging, or even fully automated, thereby significantly enhancing language accessibility for various groups. Recent progress in fine-tuning smaller language models like GPT-2 demonstrates the potential of text simplification services (Anschütz et al., 2023; Klöser et al., 2024). Furthermore, even greater quality improvements can be anticipated through the fine-tuning of LLMs, as demonstrated by the superior performance of ChatGPT in few-shot prompting compared to dedicated text simplification models (Feng et al., 2023).

However, despite their impressive capabilities, LLMs have significant flaws. Contemporary language models generate language by predicting the most likely next word, based on the parameters learned during training. LLMs were therefore trained on a vast space of textual data. As a result, the models learn social biases, hallucinate, or oversimplify complex matters with little possibility for control mechanisms (Ferrara, 2023). Moreover, they may be used intentionally to provide false information to a vulnerable group of addressees.

In this paper, we argue for the feasibility of easy-read and plain language translation or generation by LLMs and outline the potential benefits and harms induced. Thereupon, we elaborate recommendations to practitioners and developers in the context of easy-read text. More specifically, we recommend, next to ethically motivated recommendations to LLM development in general, that developers, practitioners, and the domain-specific target group should collaborate closely to minimize safety concerns and optimize the intended use of the system.

Finally, this paper raises further ethical questions for ethicists to address in future research. Who should develop text simplification systems for a vulnerable audience? Should the access to easy-read LLMs be public or restricted to a more controlled care-worker setup?

Ethical, linguistic, and computational foundations

To comprehend the impact of Natural Language Processing (NLP) and LLMs on easy-read and plain language, we will briefly introduce the ethical motivation, linguistic rules, and computational opportunities for easy-read and plain language.

A short introduction to easy-read and plain language

Natural language is inherently sophisticated and complex. Its intricacies enable us to express complex thoughts succinctly, but they can also act as barriers, excluding individuals with language comprehension difficulties from full participation in society. Plain Language and easy-read are strategies to simplify texts for different target groups (Cheung, 2017; Jones & Williams, 2017). Plain Language is a strategy to make written and spoken information easier to understand. The target audience for Plain Language includes non-native speakers, domain non-professionals, or children. Techniques used in Plain Language involve simplifying vocabulary and syntax, reducing jargon, and using familiar words in their usual context (Cutts, 2020). In contrast, easy-read takes simplification a step further, focusing on individuals with cognitive disabilities. The approach necessitates using simple words, direct speech, short sentences following the subject-predicate-object arrangement, and avoiding negations or complex tenses. By reducing language to its most basic and direct form, easy-read aims to make the information as accessible and straightforward as possible. For example, consider the plain language and easy-read translations in Table 1. In the easy-read version, the sentence structure is changed to a more straightforward subject-predicate-object arrangement, a reduced vocabulary complexity, and the use of direct speech. It also contains less precision than the original Standard English version, an often unavoidable consequence of simplification. We further used the fine-tuned text simplification models introduced by (Klöser et al., 2024) and GPT-3.5 Turbo to create the automated text simplification columns.

Despite its apparent simplicity, implementing easy-read can pose significant challenges. One obstacle is the need for more nuance. Expressing uncertain or complex situations in clear and direct statements can lead to a loss of meaning or the introduction of inaccuracies, which is particularly challenging in contexts such as legal texts where precision is crucial. Balancing the audience's needs with the integrity of the original text's meaning is delicate, requiring deep understanding and empathy for the target audience's needs (Rink, 2023).[2]

Today, there is a strong need for well-trained and educated practitioners to create easy-read and plain language texts. The task demands linguistic understanding and deep

[1] In terms of development, yet extensive computational resources are required to deploy and run LLMs.

[2] For a more comprehensive understanding of the principles of easy-read, resources such as "Making Written Information Easier to Understand for People with Learning Disabilities" by the UK Department of Health and "Guidelines for Easy-to-Read Materials" by the International Federation of Library Associations provide in-depth guidance (DHSC 2010; IFLA 2010).

Table 1 Examples of translations to plain language and easy-read

Standard Language	The President possesses the constitutional authority to veto legislation passed by Congress, thus negating the bill's effects unless it can garner a two-thirds majority support in both Houses to override the veto		
	Human-Translation	Automated Translation (GPT3.5-Turbo)[a]	Automated Translation (Fine-Tuned Model)
Plain Language	"The President can reject laws that Congress wants to pass. But if more than two-thirds of Congress agree, they can still make the law happen even if the President doesn't want it."	“The President can reject a law made by Congress. However, if enough lawmakers agree, they can still make the law happen even if the President says no.”	“The President can use a veto. This means that he says no to a law that Congress has passed. The law then does not take effect. However, the veto can be overridden if a two-thirds majority of representatives and senators vote in favor of the law The President can vote against a law.”
Easy-Read	"The President can say 'no' to new laws. But if many people in Congress say 'yes,' the law can still happen."	“The President can say no to a law made by Congress. But if most of the lawmakers agree, they can still make the law happen.”	“The President can vote against a law Then the law does not apply This is called veto power The President can only vote against laws Not against parts of laws The President can only vote against a law once Then the law does not apply Congress can then vote on the law again If 2 thirds of the people in Congress vote in favor, the law still applies Then the veto right has not worked.”

[a]Prompt: “Translate the following sentence to a) easy-read and b) plain language:”

empathy for different reader groups' abilities and needs. Achieving the inclusivity goals of plain language and easy-read without compromising the text's integrity remains a significant challenge (Chinn, 2019; Rink, 2023), Thus, the development of technical solutions, such as sophisticated language processing tools, is an active field of research.

On the feasibility of easy-read by large language models

NLP is a field of computer science and linguistics that investigates the intersection between human language and computation. Applications from NLP are automated classifications, translations, text summarizations, or text simplifications. While early systems were orchestrations of manual rule sets, machine learning became more and more influential. Recently, large deep learning models got in the focus of public discourse for their significant advances in the context of GenAI. GenAI is a branch of artificial intelligence that can generate content, such as text, images, or music, by analyzing and learning from extensive datasets, and subsequently producing outputs based on the patterns and structures it has learned during the training process. LLMs are one of the latest achievements. In their basic form, these models train to solve "fill in the blanks" exercises on vast amounts of text data extracted from the web (Devlin et al., 2019). Moreover, the pre-trained LLMs can be adapted to specialized or related tasks using traditional fine-tuning or few-shot learning. In comparison to the pre-training of an LLM or the complete training of a dedicated deep learning model, few samples suffice to adopt LLMs to reach good results on various tasks (Brown et al., 2020; Devlin et al., 2019). The example in Table 1 demonstrates that even by simple prompting (zero-shot), LLM-based chatbots can produce easy-read and plain language-like texts already.

Traditional fine-tuning

Fine-tuning denotes the specialization of, e.g., generative language models, to specialized tasks on a different dataset, e.g., text simplification (Anschütz et al., 2023). To fine-tune a model for text simplification (for instance to generate plain language), we modify the LLM to predict easier formulations instead of text completions. Research results indicate that a few thousand samples are enough to finetune models on various language problems including text simplification (Anschütz et al., 2023; Devlin et al., 2019).

Therefore, using fine-tuning may facilitate the usage of more complex NLP tasks for easy-read and plain language. By fine-tuning a translation task such as text simplification, for example, one can train an LLM to translate standard language into easy-read or plain language (cf. Table 1).[3] Further, by fine-tuning text completion tasks, one can train an LLM to generate easy-read texts based on user prompts.

In-context-learning

In contrast to the traditional fine-tuning techniques introduced in Section "Traditional fine-tuning", with the rise of GPT-3, prompting techniques offer context to an LLM to solve adapted tasks without parameter-tuning, emerged (Brown et al., 2020; Wei et al., 2022). In-context-learning refers to context prompts that give task descriptions and/or a few examples in natural language (Dong et al., 2023; Logan IV et al., 2022) or provide the model with a chain-of-thought, to solve the task (Dong et al., 2023).

Technical challenges

Next to the resulting technical opportunities, the modification of an LLM to a specific task bears several technical risks. For instance, LLMs may internalize biases from data. Natural language texts commonly reflect social biases, and presumptions, and contain explicit or implicit stereotypes (Ferrara, 2023). For instance, Wikipedia texts portray men and women differently (Wagner et al., 2015). While for example, investigating an LLM's text completion can reveal an internalized gender bias (Bhardwaj et al., 2021), finding and mitigating latent biases in a pre-trained model is not trivial (Ferrara, 2023). Moreover, domain specificity requires the simplification of text to be sensitive to different levels of background knowledge in the target group, making it particularly hard to automatically evaluate and validate the quality of the text simplification. Sufficient resources, like datasets and potentially human feedback, are essential to overcome these challenges. Especially for languages other than English, the required resources may be scarce.

The potential benefits and harms of LLM assisted easy-read

There are rarely any openly available simplification systems for most languages. Nevertheless, the research field is active, and the available resources will grow in the following years. Related areas, like text summarization, translation, or chatbots like ChatGPT, show how successful fine-tuned LLMs are. Thus, text simplification systems will most likely become widely available. We should consider the ethical

[3] (Anschütz et al., 2023) show that automatically simplified texts have characteristics of easy-read, even if incorrect and inconsistent texts still pose a problem.

aspects discussed in this section early in development, to ensure the technology's valuable effects, the potential benefits, and minimize its potential harms.

Most automation of text-generation is associated both with risks and potential benefits. The success of LLMs, in general, has led to a variety of ethical investigations of their creation, the adequacy of use-cases, the validation and robustness of their output (Kasneci et al., 2023; Kempt et al., 2023; Lund et al., 2023; Mökander et al., 2023), as well as dual-uses, misuses, and long-term consequences of automated text-generation (like fake news and misinformation (Pan et al., 2023)). All these issues are present in the production of easy-read texts as well. However, considering the different ways that easy language is produced (as a translation from a non-easy-read text to an easy-read one or as a genuinely new easy-read text, e.g.), a careful examination of the risks and benefits associated with easy language specifically is required.

The domain of easy language and plain language generation and translation by LLMs forms a specialized area in the ethics of LLMs, as the end users are mostly people from "vulnerable groups". Vulnerability is a concept that is most commonly used in research and professional ethics. The term "vulnerable group" refers to groups that are "more likely to be misled, mistreated, or otherwise taken advantage of" (Levine, 2004, p. 396). The concept was criticized in terms of its scope, a.o., for stereotyping entire groups of individuals as vulnerable (Levine et al., 2004). In the context of this article, however, we will focus on the aspect of an increased likelihood of being misled. While acknowledging that within the group of the supposed easy-read audience, the likelihood of being misled and the required language support may vary across individuals. Belonging to the part of that group, not only benefiting from but dependent on care work, presupposes a lack of language understanding capabilities and therefore qualifies for that aspect of vulnerability. To avoid stereotyping, one needs to be more careful with the prescription of vulnerability to the domain of plain language. The supposed audience of plain language does not necessarily have inherent deficits in understanding standard language: we might consider non-domain experts and laypeople more challenged by a complicated scientific text than experts; we may also consider non-native language speakers to be more challenged by a given text than native language speakers, and children to be more challenged by any text than adults. And while children may also lack the capabilities to responsibly process the information provided by LLMs and thus, may be included in the group of vulnerable users (Frenda et al., 2011), laypeople and language-learners do not. This means that easy-read and simplified texts must not only be viewed through the lens of vulnerable populations, even though they might benefit the most from LLMs learning to translate and generate this kind of text.

Instead, non-native language learners especially, as fully autonomous and independent agents who are merely missing some learnable skill, demonstrate that textual understanding is not only determined by cognitive ability but also by practical circumstances. Depending on those practical circumstances, other ethical requirements may be relevant for the responsible use of easy-read or plain-language LLM.

In the following, based on the distinctions made in Section "On the feasibility of easy-read by large language models", we examine the ethical risks and benefits that an increased use of LLMs for easy-read text might entail. These are usually understood as trade-offs that ought to be weighed against each other and require careful consideration in developing and implementing LLMs that are capable of translating or generating easy-read and plain language texts.

Translation

Turning to the use of LLMs as translators or text-simplifiers for easy-read and plain language texts, the potential benefits are rather obvious. Those who have trouble understanding long words and complex sentences are often subtly excluded from participating in public discourse and thus might be unduly disadvantaged in exercising some of their civil liberties and duties.

Automated translators for text-simplification would enable these users to live **more autonomously** as they rely less on other people's help in translating text that would otherwise remain incomprehensible to them. Users could also navigate life with **less uncertainty**, as we would expect them to understand more of the text surrounding them than they would without such a translator at hand, which also supports their autonomy as it increases decision-making capabilities. Enhancing the ability of those requiring easy-read texts for language understanding, to participate in society also constitutes a matter of **justice**: contributing to a more just society by reducing intellectual hurdles should not only be considered a desirable benefit but an obligation to realize. The more access more people have to participation in public discourse, the better (Anderson, 1999; Habermas, 1991, 1996).

Moreover, sources of **entertainment** that were previously unavailable may become accessible to the target group, not only increasing the quality of life but potentially the ability to partake in public discourse.

Lastly, given the extended availability of easy-read and plain language, care-workers may be relieved of the time-consuming task of translating texts into easy-read or plain language. Especially those who are in close or caring relationships with people who require easy-read text could direct their attention and care-work to other areas of the relationship, and thus improve the life of those depending on easy-read text from a **relational perspective**.[4]

On the other hand, the risks associated with using easy-read translators ought to be considered carefully and thoroughly. First, those who require easy-read text due to inherent difficulties with reading and understanding complex sentences and longer words should count as a **vulnerable group**, as easy-read targets an audience with cognitive or learning disabilities (Chinn & Homeyard, 2017; Sutherland & Isherwood, 2016) or other issues with processing information when presented in a complex form. Moreover, the intended audience of plain language may at least partially be categorized as vulnerable, as, e.g. children are more prone to being misled by certain information. However, also the non-vulnerable audience of plain language, e.g. non-native speakers, may be susceptible to unnoticed misunderstanding and overreliance. The audience's vulnerability and the proneness to misunderstanding, suggest the risk of **exploitation**, **abuse**, and **confusion**. Thus, some of the ethical concerns present in LLMs and especially machine-translation services are heightened due to the focus on a vulnerable audience, while other issues only emerge because of the partially vulnerable audience. While the fact that the plain-language audience might use these translators to understand crucial information about their own lives might increase their autonomy, it also increases their susceptibility to confusion and possibly avoidable harm. Similarly, human assistance in making a standard text accessible to those with comprehension issues usually has a double function: it translates texts from standard to easy-read language, but also scans for potentially harmful or exploitative text. While this process is of paternalistic nature and its absence, as we discussed earlier, would increase the autonomy of the easy-read and plain language audience, its legitimacy must be discussed in the context of a group, vulnerable to harm by manipulation and exploitation. Similar to a taster who tests food before someone else consumes it, easy-read translation made by humans can also receive a content assessment by a member of the supposed audience. The ability to spot and thus avoid scammers, for example, will be reduced if any text is automatically reproduced in easy-read text.

Relatedly, a potentially imprecise translation can cause harm and confusion, if the source text contains important information (e.g., a letter from the tax office). The lack of human control can lead to a **responsibility gap** for the potential harm caused by either imprecise, infidelitous, or unfaithful translations of easy-read text. The brittleness of LLMs in the face of adversarial attacks or prompt injections of these machines renders the robustness of appropriate output problematic (see Kempt et al., 2023).

Finally, a possible **loss of information** in translating to easy-read or plain language may occur, which can constitute different ethical concerns that should be addressed. First, considering the limits of translation to simpler language, some of the information loss can be classified as information reduction: it is not so much lost as it is intentionally reduced in complexity. For such an information reduction to be adequate, however, there must be specific guidelines about which information needs to be retained to avoid the loss of key points of the text in question. This task is made particularly daring and difficult by the fact that the adequate level of information reduction is context-dependent, e.g., on the group of addresses or the recency of the concepts used. Second, this loss may be an avoidable but unforeseeable side effect of translational LLM, as the predictability of LLM is limited by their construction. In these cases, the damage of information loss can cause harm or maintain some epistemic injustice, as the contents of the simplified text may contain harmful misrepresentations (e.g., an incorrectly translated official administrative letter, causing the receiver to lose out on benefits) or lack information that the reader should have access to (e.g., the information about the benefits someone has a claim to). Third, some limits are inherent to the algorithm and may always occur, and are thus unpreventable. The fallibility of technology ought to be kept in mind as a general point, but should not, as with other uses of fallible technology, prevent the use altogether.

Generation

So far, we have discussed the potential benefits and harms of LLM-based translations from standard language to easy-read or plain language. However, language models can not only translate a given input asked for by a user but can also produce output that functions as a response to an input, rather than a translation of such. LLMs, thus, can generate genuinely new easy-read texts.

The ability for easy-read text generation is promising to be one of the great benefits of such LLM, and can even amplify the previously outlined benefits of translational LLM. As we have seen with translator-LLM for easy-read or plain language, the ability to retrieve information that is catered to the needs of a specific group of readers enhances their capacity to navigate the world more autonomously. Requests and internet searches to answer a question of someone who prefers or requires easy-read text might be within reach now. Thus, in the context of justice, easy-read and plain language LLM may lead to the **empowerment**

[4] The point made here refers to people in a caring relationship. It should be mentioned that there are organizations that specialize in the translation of easy-read or plain language texts. While it is generally not foreseeable that those organizations will be replaced (but possibly rather augmented), we refrain from the debate on the effects on their work, as it is not directly connected to the potential benefits and harms of the supposed audience.

Table 2 Summarization of the risks and benefits of LLM based translation and generation of easy-read and plain language texts

Types of tasks	Potential benefits	Potential harms
Translation	• Increased autonomy • Relational benefits in care work • More just access to information • More just access to public discourse and partaking	• Responsibility gaps • Information loss • Exploitation & manipulation
Generation	• Amplified pot. translation benefits	• Amplified pot. translation harms • Biases • Inappropriateness

of different groups that are otherwise disadvantaged from reading complex texts, e.g., people with lower cognitive abilities, or language learners.

The real-time generation of easy-read and plain language text can have a variety of secondary positive effects, such as the previously mentioned ability to **appreciate and take part in contemporary public discourse**, and a more informed and independent public. In contrast to a translator, however, the generation would no longer require the user to first search for appropriate texts to be translated and hence, further increase the capabilities of **autonomous** participation.

The quality of text-based **entertainment** for readers of easy-read and plain language texts may also significantly improve, as these LLM may produce new fictional stories. Especially for those who are learning a new language, plain text generation can help make practicing this new language more entertaining and worthwhile.

Despite these strong reasons in favor of making easy-read and plain language LLMs widely available, we ought to caution that there are some ethical risks that should be addressed first. In the following, we will discuss how previously mentioned risks of easy-read translators are being increased, and which ones are specifically emerging with the generation of new text.

As with any LLM, the accuracy and veracity of the information provided in a given output is of chief interest and has been discussed controversially before. Hallucinations, inaccurate representations of information, and even misrepresentations are not uncommon in contemporary language models and thus we ought to expect that easy-read LLMs will be no exception (see Guerreiro et al., 2023 for an elaboration on the issue of hallucinations of LLM in machine translation). Therefore, we should consider the risk of **false information** being generated by the LLM, potentially leading to harmful outcomes. Next to this overall risk of LLMs, easy-read may introduce additional problems on this level: as we have pointed out with the information loss risk of a translation, the generation of a new text may also limit the accuracy with which a given subject matter can be portrayed while fulfilling easy-read requirements. The worry here, then lies not only in the information loss but the **loss of complexity** with which a subject is represented.

This concern is equally present in the generation of text containing information about current events. As other LLMs have shown, continuously incorporating new training data can lead to a decrease in the accuracy of the information delivered (Chen et al., 2023). The **susceptibility to misinformation** online will also translate to readers of easy-read-LLM generated content, as they might use the generation of easy-read language as their main source of information.

However, as previously stated, users of easy-read or plain language models may be more **vulnerable** to the potential misinformation presented than other users are. This impedes critical fact-checking even more, as the source of information is catered to their needs, while fact-checking sources might be in non-easy-read language yet again.

This lack of humans-in-the-loop in information generation harbors the risk of **responsibility gaps** again, as constant supervision of appropriate output is required to attribute responsibility. Both previously mentioned risks also suggest the risk of **biased output** to be taken into account (Table 2). As with other LLMs, the training data used to train an LLM often contains problematic speech or mirrors social biases, which may lead to biased outputs. Depending on the biases exhibited, this might affect vulnerable users of easy-read LLMs in their understanding of the world. It might even cause harm with a heightened risk of reinforcing social biases within the targeted audience. As other LLMs have shown the capacity for offensive statements (Neff, 2016), it stands to assume that easy-language LLMs might also produce **inappropriate, offensive, or otherwise harmful output.**

Consequences and recommendations

The deployment of LLMs for translating and generating plain language and easy-read texts ethically permissible use-cases and conditions are yet to be determined.

However, there are certain normatively motivated recommendations for the development and research in that area as a consequence of our analysis.

The potential benefits and harms outlined in Sect. "The potential benefits and harms of LLM assisted easy-read" and the fact that there are first actual implementations of text simplification models, generating and translating easy-read and plain language emphasize the need for critically reflected development.

First, on the one hand, there are consequences related to the use of LLMs for text simplification that are not unique to it but apply to the use of LLMs in general. Namely, the **mitigation of biases** and the **ensuring of appropriate** and considerably **safe text** outputs (Bender et al., 2021; Floridi, 2023; Kempt et al., 2023). However, given the increased likelihood of being misled for the vulnerable parts of the target group, the efforts in that direction should be further intensified in the validation of the system, to make its use justifiable. On the other hand, the problem of prompt injections is challenging the safety of contemporary LLMs and constitutes an active field of research (Greshake et al., 2023).

Second, in the light of **information loss, loss of complexity**, and more generally, the **understandability** of text simplification, developers encounter multiple problems. While there may be individual cases that do not satisfy the criteria sufficiently, one should strive to optimize the model concerning correctness and completeness. On the one hand, measures must be taken to avoid the unintended loss of information between standard language and simplified text. To do so, metrics from classical machine translation tasks may be adopted to validate the quality of the model. In the context of generation, the task of sound and complete texts is analog to standard language model optimization. On the other hand, there is a problem in LLM development that seems specific to text simplification models: the problem of verifying understandability. In classical LLM settings, it seems safe to assume a text to be understandable to the average user if it is similar to an average text, language-wise. Fine tunings can be made to vary across domains and levels of expertise. Especially in the context of easy-read, however, even manually written texts are typically evaluated for understandability by both easy-read experts and the potential audience. While there are rules for authors of easy-read and plain language to verify the syntactical quality of the text that may be translated to automated metrics, they have little value for the semantics of the text and do not constitute a sufficient criterion for understandability. The concepts that need to be explained in addition to the content may vary first, across groups, and second, across time. For instance, while it is safe to assume that the term "pandemic" required further explanations to some of the easy-read target groups, the concept became popular and most likely no longer needs further explanations to these groups nowadays. Thus, good performance with respect to standard quality metrics and even quality measures from the easy-read and plain language domain does not guarantee understandability to the supposed audience. We therefore recommend the close collaboration of the supposed audience, domain experts, and developers to develop highly adaptive solutions.

Moreover, given the shift of knowledge within the audience and the potential threats of information loss or loss of complexity, text simplification models should be highly auditable, allowing for a continuous evaluation of their quality.

Future research & concluding remarks

LLMs and their variety of use-cases promise to enhance the lives of many by automating time-consuming, tedious, and low-level writing tasks. Thus, as Floridi (2023) puts it, they can function as writing-assistants to elevate and alleviate challenges. This technology, in all its prowess and promise, comes with risks on different levels and for different reasons that were elaborated upon elsewhere. In this paper, we reassessed the challenges, risks, and opportunities for creating and improving LLMs for easy-read text from technical and ethical requirements. We find that easy-read and plain language LLMs can if the technical hurdles are overcome to guarantee reliability and aptness of their text translations and generations, provide valuable autonomy support on the one hand and an improvement of quality of life on the other hand. This potential to assist those who require or benefit from easy-read to navigate and understand the world ought also to be considered in the light of misuse, higher susceptibility to misinformation, and other ethical concerns that might arise. We see three specific requirements for their implementation that can fulfill the potential benefits while keeping the risks at a minimum, rendering easy-read LLMs tools that support their users' autonomy and contribute to their quality of life.

First, for those who currently live within a system that provides human caretakers, easy-read LLMs may be used as an auxiliary or complementary tool for this care work, rather than as a replacement for the care work altogether. As pointed out earlier, the limits and issues of LLMs to create inappropriate or inaccurate content are not resolved and need to be the objective of technical research in this context. The increased susceptibility to misinformation might expose users to a previously unknown level of exploitation attempts. It can, though, help caretakers to provide better and more personal care, and guide the user to a more self-secured life with a better understanding of their surroundings by an assistive tool that translates text into or generates easy-read language.

Second, we suggest that easy-read LLMs are specified for their intended uses, especially if there are no caretaker relationships present.[5] Open-domain chatbots are notoriously unsafe and thus ought to be restricted in their purpose. One can think of specialized government- or legal easy-read LLMs that translate highly complex text from administrative offices: this relieves both the text creator to also produce an easy-read option and the reader who otherwise would not understand the text. As with any technology that affects caring relationships, the just implementation of easy-read and plain language LLMs requires a careful deliberation of the effects on only on the caretaking but also on the care-giving entities. Thus, further research should be conducted on the implementation in existing social and administrative ecosystems.

Third, we contend that the question of whether these ought to be understood as goods with public access or as specialized and limited as a tool for those who need them ought to be answered. While there are some reasons in favor of keeping these easy-read LLMs limited in their access, ultimately, for proper participation in public discourse, they ought to be publicly accessible. Not only to avoid the perception of those LLMs as a crutch but also to increase the accessibility to those who may feel shame for needing this tool, or for those who cannot otherwise afford it. The goal of increasing the ability to participation in public discourse by lowering thresholds outranks other concerns. The condition for such public access to be not only permissible but morally required is that the technological concerns discussed in the previous sections ought to be resolved reliably.

If both the ethical concerns and the technical challenges can be resolved in a reliable manner, by understanding easy-read LLMs as a chance for caretakers and those taken care of to improve their quality of life as well as their lived autonomy, easy-read LLMs promise to contribute to the public good in a considerable way.

Author contributions All the authors contributed to the main idea, writing and editing of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry for Family, Senior Citizens, Women and Youth (Grant number 3923406K05).

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Does not apply.

Consent to participate Does not apply.

Consent to publish Does not apply.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Anderson, E. S. (1999). What is the point of equality? *Ethics, 109*(2), 287–337. https://doi.org/10.1086/233897

Anschütz, M., Oehms, J., Wimmer, T., Jezierski, B., & Groh, G. (2023). Language models for german text simplification: Overcoming parallel data scarcity through style-specific pre-training. https://doi.org/10.48550/ARXIV.2305.12908.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).

Bhardwaj, R., Majumder, N., & Poria, S. (2021). Investigating gender bias in BERT. *Cognitive Computation, 13*(4), 1008–1018. https://doi.org/10.1007/s12559-021-09881-2

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Jeffrey, Wu., Winter, C., Amodei, D. (2020). Language Models Are Few-Shot Learners.

Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT's behavior changing over time? *arXiv preprint* arXiv:2307.09009.

Cheung, I. W. (2017). Plain language to minimize cognitive load: A social justice perspective. *IEEE Transactions on Professional Communication, 60*(4), 448–457. https://doi.org/10.1109/TPC.2017.2759639

Chinn, D. (2019). Talking to producers of easy read health information for people with intellectual disability: Production practices, textual features, and imagined audiences. *Journal of Intellectual & Developmental Disability, 44*(4), 410–420. https://doi.org/10.3109/13668250.2019.1577640

Chinn, D., & Homeyard, C. (2017). Easy read and accessible information for people with intellectual disabilities: Is it worth it? A meta-narrative literature review. *Health Expectations, 20*(6), 1189–1200. https://doi.org/10.1111/hex.12520

Cutts, M. (2020). *Oxford guide to plain English*. Oxford University Press.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language

[5] Depending on the use-case, the classification as a medical software may make this requirement legally relevant.

Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics (pp. 4171–4186).

DHSC. (2010). Making Written Information Easier to Understand for People with Learning Disabilities. *GOV.UK*. Retrieved July 28, 2023, from https://www.gov.uk/government/publications/making-written-information-easier-to-understand-for-people-with-learning-disabilities-guidance-for-people-who-commission-or-produce-easy-read-information-revised-edition-2010.

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J., Li, L., & Sui, Z. (2023). A survey on in-context learning. *arXiv preprint* arXiv:2301.00234

Feng, Y., Qiang, J., Li, Y., Yuan, Y., & Zhu, Y. (2023). Sentence simplification via large language models. arXiv preprint arXiv:2302.11957

Ferrara, E. (2023). Should ChatGPT be biased? Challenges and risks of bias in large language models. *arXiv preprint* arXiv:2304.03738.

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(1), 15. https://doi.org/10.1007/s13347-023-00621-y

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20*(1), 20–23. https://doi.org/10.1177/0963721410396620

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023, November). Not what you've signed up for: Compromising real-world llm-integrated applications with indirect prompt injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security* (pp. 79–90)

Guerreiro, N. M., Alves, D., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). Hallucinations in large multilingual translation models. *Transactions of the Association for Computational Linguistics, 11*, 1500–1517.

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. Wiley.

IFLA. (2010). Guidelines for easy-to-read materials. Retrieved July 28, 2023, from https://ocm.iccrom.org/documents/ifla-guidelines-easy-read-materials.

Jones, N. N., & Williams, M. F. (2017). The social justice impact of plain language: A critical approach to plain-language analysis. *IEEE Transactions on Professional Communication, 60*(4), 412–429. https://doi.org/10.1109/TPC.2017.2762964

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for Good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kempt, H., Lavie, A., & Nagel, S. K. (2023). Appropriateness is all you need!. *arXiv preprint* arXiv:2304.14553.

Klöser, L., Beele, M., Schagen, J. N., & Kraft, B. (2024). German Text Simplification: Finetuning Large Language Models with Semi-Synthetic Data. *arXiv preprint* arXiv:2402.10675

Levine, C. (2004). The concept of vulnerability in disaster research. *Journal of Traumatic Stress, 17*(5), 395–402. https://doi.org/10.1023/B:JOTS.0000048952.81894.f3

Levine, C., Faden, R., Grady, C., Hammerschmidt, D., Eckenwiler, L., & Sugarman, J. (2004). The limitations of 'Vulnerability' as a protection for human research participants. *The American Journal of Bioethics, 4*(3), 44–49. https://doi.org/10.1080/15265160490497083

Logan IV, R. L., Balažević, I., Wallace, E., Petroni, F., Singh, S., & Riedel, S. (2022). Cutting Down on Prompts and Parameters: Simple Few-Shot Learning with Language Models. In *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics (pp. 2824–235).

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology, 74*(5), 570–581. https://doi.org/10.1002/asi.24750

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00289-2

Neff, G. (2016). Talking to bots: Symbiotic agency and the case of tay. *International Journal of Communication, 10*, 4915–4931.

Pan, Y., Pan, L., Chen, W., Nakov, P., Kan, M. Y., & Wang, W. Y. (2023). On the risk of misinformation pollution with large language models. *arXiv preprint* arXiv:2305.13661

Rawls, J. (1971). A theory of justice. In *Ethics: Contemporary readings*. Belknap Press/Harvard University Press.

Raz, J. (1986). *The morality of freedom*. Clarendon Press.

Rink, I. (2023). Competences for easy language translation. In S. Deilen, S. Hansen-Schirra, S. H. Garrido, C. Maaß, & A. Tardel (Eds.), *Emerging fields in easy language and accessible communication research, easy—Plain—Accessible* (pp. 231–251). Frank & Timme GmbH.

Sutherland, R. J., & Isherwood, T. (2016). The evidence for easy-read for people with intellectual disabilities: A systematic literature review. *Journal of Policy and Practice in Intellectual Disabilities, 13*(4), 297–310. https://doi.org/10.1111/jppi.12201

Wagner, C., Garcia, D., Jadidi, M., & Strohmaier, M. (2015). It's a Man's Wikipedia? Assessing gender inequality in an online Encyclopedia. *Proceedings of the International AAAI Conference on Web and Social Media, 9*(1), 454–463. https://doi.org/10.1609/icwsm.v9i1.14628

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.