Evaluating Diffusion-Based Image Generation for Easy Language Accessibility

Christoph Johannes Weber *†, Dominik Beyer †, Sylvia Rothe *
*University of Television and Film Munich, Munich, Germany
†LMU Munich, Munich, Germany
e-mail: c.weber@hff-muc.de, dominikbeyer711@gmail.com, s.rothe@hff-muc.de,

***Abstract*—Easy Language is a linguistic resource designed to facilitate comprehension for individuals with learning impairments and non-native speakers. Utilizing simplification of text, the restriction of vocabulary, and layout adjustments, Easy Language texts are constructed to ensure accessibility. Furthermore, Easy Language texts have the capacity to incorporate visual aids, such as images or symbols, to enhance comprehension. Although the use of imagery has been shown to improve understanding, it remains unclear whether visuals generated by artificial intelligence (AI) can meet the specific stylistic and semantic requirements of Easy Language. This paper investigates the potential of diffusion-based image generation to address these needs. A Stable Diffusion model was fine-tuned to produce images in a minimal, symbol-like style. Two user studies were conducted to assess the model's ability to replicate a consistent visual style and to determine whether the generated images effectively conveyed the intended meanings. The results show that participants were generally unable to distinguish AI-generated from original symbols and correctly interpreted most of the illustrated concepts. The findings suggest that diffusion models, when properly fine-tuned, are capable of producing illustrations that align with the stylistic conventions and semantic clarity required in Easy Language. However, certain abstract or emotionally nuanced concepts remain challenging to represent accurately. These results indicate that, when guided by stylistic constraints, AI-generated visuals can offer a scalable approach to producing accessible visual content.**

Keywords-image generation; accessibility; easy language.

I. Introduction

Communication is a central aspect of everyday life, whether it is necessary to coordinate daily activities, interact socially, or communicate information to broader audiences. Depending on the context and recipients, the mode and complexity of communication are adapted accordingly. However, in public domains, such as government websites, the audience often includes individuals with various cognitive and linguistic abilities. For this reason, ensuring broad accessibility becomes a key concern.

An approach to making written content more accessible is *Easy Language*, a simplified form of communication designed primarily for people with cognitive or learning difficulties. It is governed by a set of formal rules that emphasize short sentences, familiar vocabulary, and the use of supportive visuals, such as symbols or images [1][2]. According to the cognitive theory of multimedia learning, the combination of verbal and visual information can improve comprehension and reduce cognitive load [3]. However, the visuals used in Easy Language must adhere to strict stylistic conventions: they must be consistent in style, placed near the corresponding text, and avoid redundancy or ambiguity [1].

In practice, creating these images is a complex and iterative process. Translators usually begin by drafting a visual concept for a specific word or phrase. This is given to a designer who produces an initial sketch that is reviewed by a test group, often people with learning difficulties. The image is reviewed multiple times based on user feedback until the intended meaning is clearly understood [4]. While this process ensures clarity, it is time-consuming and may limit the timely dissemination of accessible information.

Previous research has examined the role of visuals in Easy Language using methods, such as comprehension tests [5], eye-tracking [6], and reading speed analysis [7]. These studies have also evaluated different visual formats, from realistic photographs to symbolic representations [8]. However, the potential role of AI in automating this process has not yet been explored.

Recent advances in AI-based image generation, particularly diffusion models, such as Stable Diffusion [9], DALL-E [10][11] or Midjourney [12] have shown that machines can generate visually coherent and stylistically adaptive images based on text prompts. These models can be further fine-tuned to reflect specific visual styles, making them promising candidates for generating accessible visuals. Previous work has shown that some AI-generated images are indistinguishable from real images [13], but their applicability in accessibility contexts, such as Easy Language remains unclear.

This paper investigates whether AI-generated images, produced via a fine-tuned Stable Diffusion model, can support Easy Language communication. Specifically, we assess (1) whether the model can reproduce a consistent visual style aligned with existing symbol sets, and (2) whether the generated images are expressive enough to unambiguously convey intended meanings.

To this end, a Stable Diffusion model was fine-tuned using Low-Rank Adaptation (LoRA) [14] on a minimalist black-and-white pictogram data set derived from Picto-Selector [15]. Two user studies were conducted: one to test style fidelity, the other to evaluate semantic expressiveness. Participants were generally unable to distinguish the AI-generated images from originals and correctly interpreted most of the illustrated concepts.

Our findings suggest that AI-generated imagery, when guided by specific stylistic constraints, can support the goals of Easy Language and may offer a scalable alternative to manual illustration processes.

The remainder of this paper is structured as follows: Section II introduces Easy Language. Section III reviews related work. Section IV describes the model fine-tuning. Section V

presents two user studies. Section VI reports the findings. Section VIII discusses future work and concludes.

II. Background

Easy Language emerged as a means to promote inclusion and equal access to information, especially for people with cognitive or learning difficulties. Originating in the 1960s and gaining traction in Germany in the 1990s, it encompasses simplified versions of the standard language [16]. It reduces linguistic complexity through short sentences, simple vocabulary, minimal use of connectors, and the inclusion of images and adjusted layouts [1]. Terms, such as easy-to-read, clear language, or simplified language are often used interchangeably across countries. "Leichte Sprache" is the German adaptation, with this work focusing on its regulations [2][17]. Easy Language is distinct from Plain Language, which is less formalized and aims to reduce stigmatization. An intermediate form, "Easy Language Plus", has also been proposed [18].

Target Groups: Easy Language serves a diverse population, including people with learning disabilities, low literacy, sensory impairments, or those affected by migration [16][19]. However, its implementation varies between countries. While often developed for people with intellectual disabilities, it also benefits those facing temporary or situational communication barriers. Despite its accessibility goals, Easy Language sometimes faces resistance due to its distinct appearance or simplified style, potentially leading to stigmatization or rejection by target users [18]. Therefore, texts should be neutral in tone and format and provided across various media formats, not just online.

Guidelines: Rulebooks for Easy Language differ by country and context. In Germany, the *Netzwerk Leichte Sprache e.V.* and *Duden Leichte Sprache* offer key guidance [2][17]. These include linguistic, textual, and visual rules, often influenced by international frameworks, such as *Inclusion Europe* [20] or the *International Organization for Standardization* (ISO) [21]. Despite wide application, many of these rules lack a scientific basis. Current regulations, such as those defined in the *German web accessibility regulation* (BITV) [22] and in emerging national standards from the German Institute for Standardization (DIN) [23], govern accessibility on official websites. However, inconsistencies and vague formulations in these rulebooks challenge objective evaluation and implementation.

Research: Empirical research on Easy Language remains limited but is growing. Existing studies have evaluated rules through text simplification, word frequency, or visual aids [19][24], but many current guidelines are based on expert opinion rather than data. Research centers, such as the University of Hildesheim, are working to establish a scientific foundation [25]. Three main areas are being explored: text production, user perception, and translation practices [26]. Interdisciplinary perspectives also examine social and economic dimensions.

Image Support: Visuals, such as symbols and photographs, are widely used to support comprehension in Easy Language, particularly for individuals with cognitive or learning difficulties [5][27][8]. Symbols can vary in clarity, ranging from easily guessable (transparent) to learnable (translucent) and abstract (opaque) [5]. As illustrated in Figure 1, these categories reflect different levels of intuitiveness. Transparent symbols, for example a simple drawing of a dog to represent the word "dog", are generally the most effective, especially for individuals with intellectual disabilities [28][29]. Photographs are often seen as more transparent than symbols because they resemble real-life objects and actions more closely. Studies suggest they may be more effective in supporting comprehension, particularly for abstract or complex concepts [30][8]. However, both symbols and photographs can cause misinterpretation if poorly chosen or overly abstract. Their effectiveness depends on factors, such as user familiarity, visual processing ability, and attention span [31]. The cognitive theory of multimedia learning supports the use of visuals, stating that multi-modal information presentation reduces cognitive load and improves understanding [3]. However, visuals can also lead to overload or hinder comprehension if used excessively or without context [32][19]. Research findings remain mixed [24], highlighting the need to apply images with careful consideration of the target audience and communicative intent.

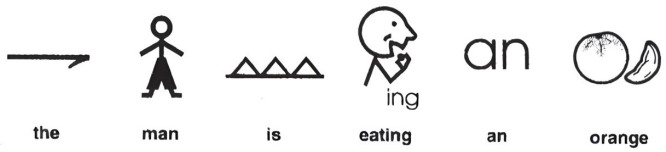

Figure 1. Illustration of opaque (“the”, “is”), translucent (“eating”), and transparent (“man”, “orange”) symbols used to support sentence comprehension [5].

III. Related Work

Several studies have examined whether humans can distinguish AI-generated images from real photographs. Lu et al. [13] conducted a large-scale study using Midjourney-generated images and internet photos, finding that participants correctly identified real images 66.9% of the time and AI-generated ones 55.8% of the time. Notably, prior exposure to AI-generated content improved classification accuracy, and images featuring people were easier to assess than object-based images. Gal et al. [33] and Ruiz et al. [34] compared different fine-tuning methods for diffusion models. While Ruiz et al. found DreamBooth to outperform Textual Inversion in terms of fidelity and subject likeness, these studies did not include real images or measure human ability to distinguish image sources.

Research into symbol comprehension has highlighted the role of visual resemblance and familiarity. Mirenda et al. [29] showed that symbols more closely resembling real objects were easier for non-speaking individuals with cognitive impairments to interpret. Similar findings were reported by Bloomberg et al. [35], who ranked symbol sets based on participant ratings. Dada et al. [36] demonstrated that children with mild intellectual disabilities could successfully match high-iconicity symbols with labels. Hartley et al. [37] emphasized that colored images improved symbolic understanding in children with autism.

Schlosser et al. [38] found that animated symbols enhanced interpretability, particularly when paired with text.

Research on the effectiveness of image-supported Easy Language remains limited and somewhat inconsistent. Rivero-Contreras et al. [6] used eye-tracking to study dyslexic readers and found that simplified text and illustrative support both contributed to improved processing. Jones et al. [27] reported improved comprehension in adults with learning disabilities when symbols were placed above individual words. Noll et al. [8] found that photo-supported Easy Language enhanced performance in mathematical tasks for students with and without special needs, whereas symbols showed no such effect, which suggests that the benefits may depend on the specific task. Poncelas and Murphy [5] observed no immediate benefit from symbols in manifestos but noted improved comprehension after repeated exposure, which highlights the importance of symbol familiarity. Conversely, Hurtado et al. [28] compared Easy Language leaflets with and without images and found no significant difference in information retention. Similarly, Parsons and Sherwood [39] implemented Widgit symbols in legal information leaflets for detainees with learning disabilities but relied solely on stakeholder satisfaction rather than comprehension metrics. Cardone [40] questioned the reliability of visual-based questioning methods and cautioned against assuming pictures always enhance understanding. A more recent user study systematically evaluated how well different AI-generated images illustrate simplified texts for accessibility purposes [41]. Involving participants from the target group, the study found that while visual fidelity was often high, semantic clarity varied greatly depending on the model and prompt formulation. These findings underline the need for human-in-the-loop approaches when using AI imagery in Easy Language.

Overall, while some studies support the potential of image support in Easy Language, results are mixed and highly dependent on task, audience, and design choices. This study extends prior work by introducing AI-generated imagery as a new visual support modality for Easy Language.

IV. Implementation

Data Set: The dataset used for fine-tuning was sourced from the Picto-Selector application, which contains over 34,000 pictograms (see Figure 2) for creating visual schedules [15]. Specifically, the Pictogenda symbol set was used due to its consistent black-and-white, minimalist visual style. This style is visually similar to Widgit symbols [42], which have been successfully used in Easy Language contexts. From the original set of 420 symbols, a total of 99 images were selected for fine-tuning. The reduction was based on two main criteria: (1) stylistic consistency, as images that diverged visually from the core set were excluded, and (2) prompt suitability, as symbols with overly complex content, such as overlapping objects or difficult pose, could not be described effectively in simple prompts. Since each image is accompanied by a textual prompt during training, inadequate or ambiguous prompts could compromise learning quality. The final set reflects a balance between visual homogeneity and prompt clarity. Each image was manually captioned to guide training, using a structured prompt format that included a unique style identifier ("pl41nl4ng"), a detailed description of the image, and stylistic tags (e.g., “black background”).

Figure 2. Sample images from the Pictogenda data set [15] used for fine-tuning. From left to right: (1) person waving, (2) wrapped gift box, (3) dog, (4) coastal scene with lighthouse and sailboat, (5) two people standing together. Each image is rendered in a simplified black-and-white style.

Fine-Tuning: Stable Diffusion v1.5 [43], trained on LAION-Aesthetics v2 5+ [44], was selected as the base model due to its open-source availability and robust latent diffusion architecture [9]. A fine-tuned autoencoder [45] was used for latent space transformations. LoRA was applied to fine-tune both the U-Net and the text encoder with reduced parameter overhead. DreamBooth’s prior preservation [34] was not necessary due to the single-style training objective.

The U-Net was trained with a learning rate of 5e-6, the text encoder with 2.5e-6, and a rank of 256. AdamW8bit optimization was used. Training ran for 29,700 steps over 30 epochs, with each image used ten times per epoch on a Tesla T4 GPU. During training, one image per epoch was sampled to monitor the model’s progress. Of the 30 total epochs, images from the first six were discarded due to low visual quality, leaving 24 candidate models for evaluation. The full training configuration, model weights, and dataset are available on Hugging Face [46]. The training script was based on an adapted notebook [47], using the kohya-trainer repository [48].

Image Generation: Images were generated via the *Stable Diffusion Web UI* implementation for Google Colab [49][50], using the same base model, LoRA weights, and autoencoder as during training. Prompts closely followed the training captions. A consistent negative prompt containing terms like “deformed” or “bad art” was used to suppress undesired output, alongside three quality-enhancing embeddings [51][52][53]. Most images were generated with 30 diffusion steps, the “Euler a” sampler, a Classifier-Free Guidance (CFG) scale of 8, and a resolution of 512×512 pixels. ControlNet [54] was used to control human poses and ensure structural consistency across styles.

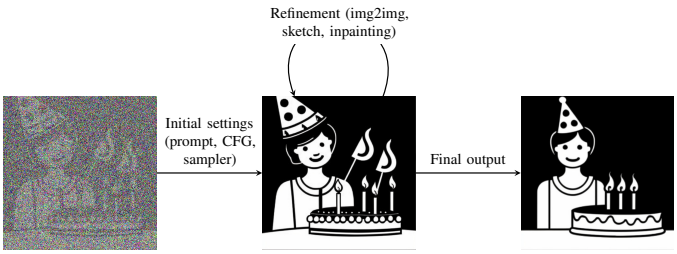

Figure 3. Illustration of the workflow for generating the images used in the user studies. An initial image was created using text2img. The image was then refined, e.g., with img2img. The final output was used in the user studies.

Since initial outputs were often imperfect, a multi-step refinement process followed. Sketching was used to remove elements, inpainting to add or replace content, and img2img to enhance visual quality. These were applied with denoising strengths between 0.1 and 0.9. To generate both black and white background versions, ControlNet was used to transfer structure, and prompts were adjusted accordingly. The full generation workflow is shown in Figure 3, and all final images (and their training counterparts) are included in Appendix A Figures 6–24.

V. User Studies

To evaluate the suitability of AI-generated images for Easy Language, two user studies were conducted. The first study examined whether diffusion models can replicate a consistent visual style, while the second explored whether the generated images are expressive enough to clearly convey specific meanings.

A. Study 1: Visual Distinctiveness

The first study assessed whether participants could distinguish between AI-generated images and original pictograms from the Pictogenda set [15]. Since Easy Language materials should maintain a consistent image style throughout [1], it is important to determine whether fine-tuned diffusion models can replicate a given visual aesthetic. This is particularly relevant because conventional diffusion models are unlikely to be familiar with Easy Language imagery.

Fifteen participants, mostly computer science students, took part in this web-based study. The experiment was conducted via an online survey platform and presented in two parts. In part one, participants were shown 40 images: 20 generated by the fine-tuned diffusion model and 20 original Pictogenda icons. Each generated image had a content-matched original counterpart. The image sets were equally divided between transparent and translucent symbols, following classification schemes, such as those by Poncelas and Murphy [5]. Participants were asked to judge whether each image was AI-generated or not. In part two, image pairs were shown again and participants had to select which image better illustrated a given concept, or mark both as equally good, similar to the setup used by Lu et al. [13].

Before the task, participants viewed a short introduction and example images to become familiar with the Pictogenda style. Their judgments were recorded, and optional text input fields captured the reasoning behind classification choices.

B. Study 2: Expressiveness for Easy Language

The second study evaluated whether AI-generated images could effectively convey meanings to people with cognitive impairments, a key requirement for Easy Language illustrations. A total of 42 participants took part, many of whom were members of the *Netzwerk Leichte Sprache e.V.* [17]. The group included both Easy Language translators and testers, some with learning difficulties.

The study was designed in consultation with an Easy Language expert and implemented using the same online survey platform. To reduce cognitive load, only 20 images were shown, half transparent, half translucent, and all instructions were written and verified in Easy Language. Images were displayed with white backgrounds, based on accessibility recommendations from the translator. Participants were asked to type what they thought each image meant, using simple text responses limited to 100 characters. To reduce frustration and mitigate the risk of participants guessing, all answers were optional.

To evaluate the answers, a three-level scoring system was applied: correct (1), partially correct (0.5), and incorrect (0). Blank responses were scored as incorrect. Partially correct answers either correctly described the image without identifying the intended meaning or mixed correct and incorrect concepts.

VI. Results

A. Fine-Tuning Stable Diffusion

To determine the best model checkpoint, one image was generated after each of the 30 training epochs. The first six epochs were excluded due to insufficient visual quality, leaving 24 candidate models. For each candidate, 10 images were generated with LoRA strength values between 0.0 and 1.0 in steps of 0.1, resulting in a total of 240 images (see an excerpt in Figure 4). All generations used the same prompt, seed, and settings to allow for consistent comparison. The final model, from epoch 19 with a strength value of 0.8, was selected based on visual evaluation. It produced images with correct composition, no unintended features such as eyebrows or detailed fingers, and strong alignment with the input prompt.

B. Visual Distinctiveness Study

This study investigated whether participants could distinguish between AI-generated and original images. Among 15 participants, the overall classification accuracy was 47.7%, which is not significantly better than random guessing. Prior experience with AI-generated images had no significant effect on performance. Filtering for image quality (e.g., blurred lines) or focusing on human-centered images also yielded no improvements. These findings are consistent with Lu et al. [13], who reported that participants struggled to distinguish AI-generated from real images, although their study suggested that human-centered imagery may be slightly easier to classify, which was not confirmed here. Participants reported relying on features, such as line sharpness, object proportions, facial expressions, and finger details in their decision-making. However, these cues did not result in reliable classification. In a follow-up task, participants compared image pairs (AI vs. original) and indicated which better conveyed the intended concept. Across 20 comparisons, AI-generated images were preferred 120 times, while original images were chosen 31 times; 149 comparisons were rated as equally good.

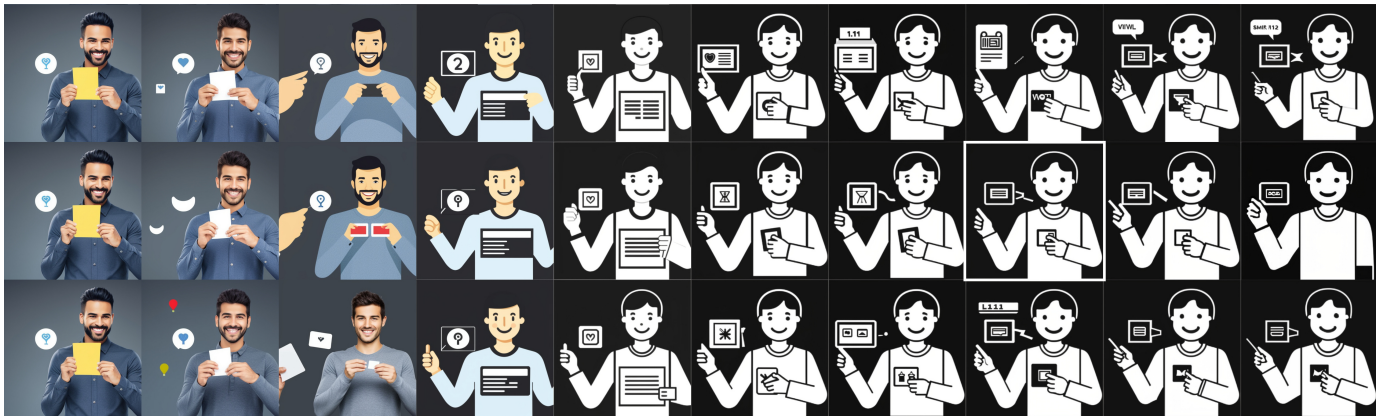

Figure 4. An overview of the training results: Rows represent epochs, columns represent influence strength.

C. Evaluation of AI-Generated Images for Easy Language

In the second study, 42 participants, including Easy Language users and translators, were asked to describe the meaning of 20 AI-generated images. The overall mean accuracy (acc) was high (acc = 0.898), with transparent images recognized significantly better (acc = 0.946) than translucent ones (acc = 0.851). This difference was confirmed by a Wilcoxon signed-rank test ($p < 0.001$). A concept-level analysis showed that 17 out of 20 concepts had a mean accuracy above 0.85. *Fish* and *sad* were recognized perfectly by all participants (Figures 15 and 23). *Headache* showed the lowest recognition rate with a mean accuracy of (acc = 0.367), followed by *coffee* (acc = 0.756) and *angry* (acc = 0.767) (Figures 16, 12, and 7).

To assess how consistently participants interpreted each concept, we calculated mean accuracy scores and 95% confidence intervals. These help evaluate the reliability of the results and highlight differences in interpretive agreement (see Figure 5).

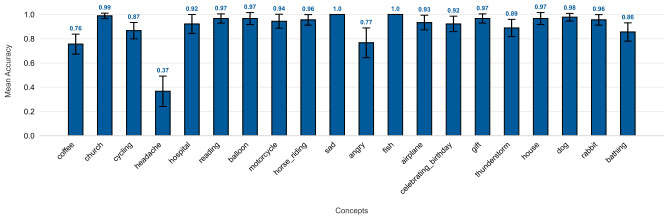

Figure 5. Mean accuracy per concept with 95% confidence intervals across all participants.

A Friedman test ($p < 0.001$) revealed significant differences between concepts. A Nemenyi post hoc test further showed that *headache* differed significantly from nearly all other concepts, and *coffee* differed from *church*, *fish*, and *sad* (Figures 16, 12, 11, 15, and 23). To identify differences between high and low performers, the 25% bottom and 25% top participants were compared using Mann–Whitney U tests. Significant differences were found for *headache*, *angry*, *thunderstorm*, *cycling*, and *hospital*, with lower scores in the bottom group for all these concepts (Figures 24, 13, and 17). To further examine individual variation among lower-performing participants, the 50% and 25% with the lowest mean accuracy were analyzed separately (n = 21 and n = 11). Both groups still achieved relatively high average scores (acc = 0.830) and (acc = 0.777), but showed more variation between concepts. In the 50% group, *headache* differed significantly from 14 other concepts. In the 25% group, although the Friedman test remained significant ($p < 0.001$), the Nemenyi test showed no pairwise differences, likely due to the small sample. Participants also provided free-text feedback. One noted that some images appeared too childlike, while another questioned the necessity of visualizing certain concepts at all.

Overall, results indicate that fine-tuned diffusion models can generate images that are visually consistent and semantically meaningful enough to support Easy Language, though with clear limitations for certain abstract or emotionally nuanced concepts.

VII. Discussion

The overarching goal of this work was to investigate whether AI-generated images can support Easy Language. Two main conditions were examined: first, whether diffusion models can replicate the visual style of existing image sets, and second, whether they can generate images that are expressive and unambiguous enough to convey meaning. These goals were addressed through two user studies.

The visual style used in the studies was not explicitly designed for Easy Language but was selected based on expert recommendation and its similarity to successful sets like *Widgit* [42]. While this choice introduces a potential limitation, the focus was on the model's ability to replicate and express a given style consistently. Some participants commented that the tested concepts were too simple or did not require visualization. However, widely recognizable concepts were chosen deliberately to isolate image expressiveness from concept familiarity. Although the studies reused some concepts seen during training, the prompts and generation settings were altered, reducing the risk of overfitting.

In the first study on visual distinctiveness, participants were unable to reliably distinguish between AI-generated and non-AI-generated images. Mean accuracy remained around 0.5, indicating chance-level performance. No significant difference was observed between participants with or without prior

experience with AI-generated images. These findings suggest that diffusion models can successfully mimic the style of existing image sets using relatively small datasets (~100 images). However, high variance in individual performance and the small sample size (15 participants) make it difficult to draw general conclusions. A larger sample is needed to further investigate whether certain groups are more capable of this task. Many participants reported relying on image sharpness as a distinguishing factor, as AI-generated images were consistently sharp while non-AI images varied in clarity. Although blurred images were excluded from part of the analysis, results remained non-significant. This supports prior findings by Lu et al. [13], who also observed that participants struggled to identify AI-generated images, especially for people-centered content.

The second study, focused on expressiveness, showed that AI-generated images achieved high recognition accuracy (acc = 0.898). Transparent images were understood more reliably than translucent ones, mostly due to low scores on a few specific concepts like *headache* and *angry*, which involve emotional or abstract meaning. Participants with lower overall scores particularly struggled with these concepts, suggesting that emotional content remains a challenge for current diffusion models. At the same time, concepts like *sad* were correctly identified by all participants, underscoring the importance of content and design choices. Since the study included translators and experts rather than exclusively people with learning difficulties, the results provide only limited insight into Easy Language's target audience. However, subgroup analysis of lower-performing participants offered useful indications of where comprehension breaks down and where image refinement may be needed.

Overall, the fine-tuned diffusion model was effective in reproducing the visual style and conveying meaning for a majority of tested concepts. Abstract or emotional concepts proved more difficult, highlighting the importance of iterative testing with target users. While AI-generated images show promise for supporting Easy Language, human feedback remains essential. In practice, generating effective images often requires additional tools, such as ControlNet [54], in combination with text prompts, which may limit accessibility for non-technical users, such as Easy Language translators.

VIII. Conclusion and Future Work

AI-generated images show significant potential for supporting Easy Language by providing scalable, on-demand visual content tailored to accessibility needs. This study evaluated whether a fine-tuned diffusion model can (1) replicate the coherent visual style required for Easy Language and (2) produce illustrations that clearly convey intended meanings. A Stable Diffusion model was fine-tuned using the LoRA method [14] on a minimalist pictogram dataset derived from Picto-Selector [15], and tested in two user studies. Results indicate that participants were generally unable to distinguish the AI-generated images from original symbols [13], and that most generated images were interpreted correctly, with an overall accuracy of nearly 90%. This suggests that diffusion models can effectively support the visual dimension of Easy Language communication. However, challenges remain for abstract or emotional concepts, such as *headache* or *angry*, particularly among lower-performing participants. These limitations highlight the need for more expressive and semantically aware generation techniques. Interpretation of the findings must consider certain constraints: the visual style was deliberately minimalist, and the participant pool, while diverse, was small and only partially representative of the Easy Language target audience. Generalizing the results to other styles or broader user groups requires further validation. Importantly, the study shows that with appropriate fine-tuning, diffusion models can produce illustrations that align with key stylistic and semantic expectations in Easy Language contexts. This balance of visual consistency and conceptual clarity is essential for accessible communication and positions AI-generated imagery as a viable component in inclusive design workflows.

Future work should build on these findings by exploring several open directions. One area of interest is whether the lower recognition rates observed among certain user subgroups (e.g., the lowest-performing 25%) are due to limitations in the model or to user-related factors such as concept familiarity or cognitive load. Dedicated studies focusing on individuals from the Easy Language target group could yield deeper insights. The current findings are specific to a minimalist visual style. It remains unclear whether diffusion models can maintain semantic clarity and stylistic consistency when applied to more complex or detailed styles. Comparative studies involving varying visual styles would help assess the generalizability of these results. Additionally, testing whether humans can still distinguish AI-generated from non-AI images in more detailed styles would be valuable. To improve the robustness of the findings, future research should increase the sample size and diversity of tested concepts. A broader participant pool and concept range could help validate the expressiveness and stylistic fidelity of AI-generated images more comprehensively. Recent developments in controllable image generation, such as ControlNet, StyleAlign, or prompt-to-prompt editing, could further enhance the expressiveness and clarity of visuals in accessibility contexts [54]. These tools offer fine-grained control over layout, pose, and visual style, making them promising for generating context-aware illustrations in Easy Language workflows. Finally, integrating user-driven image generation into Easy Language workflows may enable adaptive support. By allowing users to highlight parts of a text and generate matching illustrations, accessibility could become more personalized. However, this requires robust text-to-prompt models that can convert vague or minimal text into meaningful image prompts, an area where current text-based guidance still has limitations. To fully realize this potential, image generation must remain embedded in iterative, human-centered design processes. As controllable generation tools continue to evolve, they offer new opportunities for generating personalized and context-sensitive visual supports for diverse user needs.

REFERENCES

[1] U. Bredel and C. Maaß, *Easy Language: Theoretical Foundations and Practical Guidance*. Berlin: Dudenverlag, 2016, Original title: *Leichte Sprache: theoretische Grundlagen – Orientierung für die Praxis*, ISBN: 3411756160.

[2] C. Maaß, *Easy Language: The Rulebook*. Münster: Lit-Verlag, 2015, Original title: *Leichte Sprache. Das Regelbuch*, ISBN: 978-3-643-12907-9.

[3] R. E. Mayer, "Cognitive theory of multimedia learning," in *The Cambridge Handbook of Multimedia Learning*, ser. Cambridge Handbooks in Psychology, R. E. Mayer, Ed., 2nd ed., Cambridge: Cambridge University Press, 2014, pp. 43–71. DOI: 10.1017/CBO9781139547369.005.

[4] *Lebenshilfe for people with intellectual disabilities bremen (registered association)*, https://www.leichte-sprache.de/, Accessed: 2025-05-31.

[5] A. Poncelas and G. Murphy, "Accessible information for people with intellectual disabilities: Do symbols really help?" *Journal of Applied Research in Intellectual Disabilities*, vol. 20, pp. 466–474, 2007. DOI: 10.1111/j.1468-3148.2006.00334.x.

[6] M. Rivero-Contreras, P. Engelhardt, and D. Saldaña, "An experimental eye-tracking study of text adaptation for readers with dyslexia: Effects of visual support and word frequency," *Annals of Dyslexia*, vol. 71, pp. 1–18, 2021. DOI: 10.1007/s11881-021-00217-1.

[7] S. Schmutz, A. Sonderegger, and J. Sauer, "Easy-to-read language in disability-friendly web sites: Effects on nondisabled users," *Applied Ergonomics*, vol. 74, pp. 97–106, 2019. DOI: 10.1016/j.apergo.2018.08.013.

[8] A. Noll, J. Roth, and M. Scholz, "Overcoming reading barriers in inclusive mathematics education – a comparative study of visual and linguistic support measures," *Journal für Mathematik-Didaktik*, vol. 41, pp. 157–190, 2020, Original title: *Lesebarrieren im inklusiven Mathematikunterricht überwinden – visuelle und sprachliche Unterstützungsmaßnahmen im empirischen Vergleich*. DOI: 10.1007/s13138-020-00158-z.

[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv: 2112.10752 [cs.CV].

[10] A. Ramesh *et al.*, *Zero-shot text-to-image generation*, 2021. arXiv: 2102.12092 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2102.12092.

[11] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, *Hierarchical text-conditional image generation with clip latents*, 2022. arXiv: 2204.06125 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2204.06125.

[12] *Midjourney*, https://www.midjourney.com/, Accessed: 2025-05-31.

[13] Z. Lu *et al.*, *Seeing is not always believing: Benchmarking human and model perception of ai-generated images*, 2023. arXiv: 2304.13023 [cs.AI].

[14] E. J. Hu *et al.*, *Lora: Low-rank adaptation of large language models*, 2021. arXiv: 2106.09685 [cs.CL].

[15] *Picto-selector*, https://www.pictoselector.eu/, Accessed: 2025-05-31.

[16] C. Lindholm and U. Vanhatalo, *Handbook of Easy Languages in Europe* (Easy - Plain - Accessible). Berlin: Frank & Timme, 2021, ISBN: 9783732907717.

[17] *Netzwerk leichte sprache e.v.* https://www.leichte-sprache.org/, Accessed: 2025-05-31.

[18] S. Hansen-Schirra and C. Maaß, *Easy Language - Plain Language - Easy Language Plus: Perspectives on Comprehensibility and Stigmatisation* (Easy – Plain – Accessible). Berlin: Frank & Timme, 2020, ISBN: 9783732906918.

[19] M. González-Sordé and A. Matamala, "Empirical evaluation of easy language recommendations: A systematic literature review from journal research in catalan, english, and spanish," *Universal Access in the Information Society*, 2023. DOI: 10.1007/s10209-023-00975-2.

[20] *Inclusion europe*, https://www.inclusion-europe.eu/, Accessed: 2025-05-31.

[21] I. J. 1. 35, "Information technology – user interfaces – requirements and recommendations on making written text easy to read and understand," International Organization for Standardization, Standard ISO/IEC 23859:2023, 2023.

[22] *Regulation for the creation of accessible information technology (bitv 2.0)*, https://www.bmas.de/DE/Service/Gesetze-und-Gesetzesvorhaben/barrierefreie-informationstechnik-verordnung-2-0.html, Original title: *Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (BITV 2.0)*. Accessed: 2025-05-31, 2023.

[23] D. e. V., "E din spec 33429:2023-04 "recommendations for german easy language"," DIN Deutsches Institut für Normung e. V., Standard E DIN SPEC 33429:2023-04, 2023, Original title: *E DIN SPEC 33429:2023-04 "Empfehlungen für Deutsche Leichte Sprache"*.

[24] I. Fajardo *et al.*, "Easy-to-read texts for students with intellectual disability: Linguistic factors affecting comprehension," *Journal of applied research in intellectual disabilities : JARID*, vol. 27, 2013. DOI: 10.1111/jar.12065.

[25] C. Maaß, I. Rink, and S. Hansen-Schirra, "Easy language in germany," *Handbook of Easy Languages in Europe*, p. 191, 2021. DOI: 10.26530/20.500.12657/52628.

[26] S. Hansen-Schirra and C. Maaß, *Easy Language Research: Text and User Perspectives* (Easy – Plain – Accessible). Berlin: Frank & Timme, 2020, ISBN: 978-3-7329-0688-8.

[27] F. Jones, K. Long, and W. Finlay, "Symbols can improve the reading comprehension of adults with learning disabilities," *Journal of intellectual disability research : JIDR*, vol. 51, pp. 545–550, 2007. DOI: 10.1111/j.1365-2788.2006.00926.x.

[28] B. Hurtado, L. Jones, and F. Burniston, "Is easy read information really easier to read?" *Journal of intellectual disability research : JIDR*, vol. 58(9), pp. 822–829, 2013. DOI: 10.1111/jir.12097.

[29] P. Mirenda and P. A. Locke, "A comparison of symbol transparency in nonspeaking persons with intellectual disabilities," *The Journal of speech and hearing disorders*, vol. 54(2), pp. 131–140, 1989. DOI: 10.1044/jshd.5402.131.

[30] R. Sutherland and T. Isherwood, "The evidence for easy-read for people with intellectual disabilities: A systematic literature review: The evidence for easy-read for people with intellectual disabilities," *Journal of Policy and Practice in Intellectual Disabilities*, vol. 13, 2016. DOI: 10.1111/jppi.12201.

[31] P. Mirenda, "Designing pictorial communication systems for physically able-bodied students with severe handicaps," *Augmentative and Alternative Communication*, vol. 1, pp. 58–64, 1985. DOI: 10.1080/07434618512331273541.

[32] J. Sweller, J. J. G. Van Merrienboer, and F. Paas, "Cognitive architecture and instructional design," *Educational Psychology Review*, vol. 10, 1998. DOI: 10.1023/a:1022193728205.

[33] R. Gal *et al.*, *An image is worth one word: Personalizing text-to-image generation using textual inversion*, 2022. arXiv: 2208.01618 [cs.CV].

[34] N. Ruiz *et al.*, *Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation*, 2023. arXiv: 2208.12242 [cs.CV].

[35] K. Bloomberg, G. R. Karlan, and L. L. Lloyd, "The comparative translucency of initial lexical items represented in five graphic symbol systems and sets," *Journal of speech and hearing research*, vol. 33(4), pp. 717–725, 1990. DOI: 10.1044/jshr.3304.717.

[36] S. Dada, A. Huguet, and J. Bornman, "The iconicity of picture communication symbols for children with english additional language and mild intellectual disability," *Augmentative and alternative communication (Baltimore, Md. : 1985)*, vol. 29, pp. 360–373, 2013. DOI: 10.3109/07434618.2013.849753.

[37] C. Hartley and M. Allen, "Symbolic understanding of pictures in low-functioning children with autism: The effects of iconicity and naming," *Journal of autism and developmental disorders*, vol. 45, pp. 15–30, 2013. DOI: 10.1007/s10803-013-2007-4.

[38] R. Schlosser *et al.*, "Animation of graphic symbols representing verbs and prepositions: Effects on transparency, name agreement, and identification," *Journal of speech, language, and hearing research : JSLHR*, vol. 55, pp. 342–58, 2011. DOI: 10.1044/1092-4388(2011/10-0164).

[39] S. Parsons and G. Sherwood, "A pilot evaluation of using symbol-based information in police custody," *British Journal of Learning Disabilities*, 2015. DOI: 10.1111/bld.12140.

[40] D. Cardone, "Exploring the use of question methods: Pictures do not always help people with learning disabilities," *The British Journal of Development Disabilities*, vol. 45, no. 89, pp. 93–98, 1999. DOI: 10.1179/096979599799155894.

[41] M. Anschütz, T. Sylaj, and G. Groh, *Images speak volumes: User-centric assessment of image generation for accessible communication*, 2024. arXiv: 2410.03430 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2410.03430.

[42] *Widgit symbols*, https://www.widgit.com/, Accessed: 2025-05-31.

[43] *Stable diffusion v1.5*, https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5, Accessed: 2025-05-31.

[44] C. Schuhmann, *Laion-aesthetics v2 5+*, https://web.archive.org/web/20230331034058/https://huggingface.co/datasets/ChristophSchuhmann/improved_aesthetics_5plus, Accessed: 2025-05-31.

[45] S. AI, *Sd-vae-ft-mse-original autoencoder*, https://huggingface.co/stabilityai/sd-vae-ft-mse-original, Accessed: 2025-05-31.

[46] Hugging Face, *PlainLang Collection*, https://huggingface.co/collections/bomdey/plainlang-65663ad0f450504854ce6145, Accessed: 2025-05-31, 2025.

[47] F. Taqwa, *Fine-tuning notebook*, https://colab.research.google.com/github/Linaqruf/kohya-trainer/blob/main/kohya-LoRA-dreambooth.ipynb, Accessed: 2025-05-31.

[48] F. Taqwa, *Fine-tuning repository*, https://github.com/Linaqruf/kohya-trainer/tree/main, Commit: 3d494d8.

[49] *Stable diffusion web ui colab*, https://github.com/camenduru/stable-diffusion-webui-colab, Accessed: 2025-05-31.

[50] *Stable diffusion web ui*, https://github.com/AUTOMATIC1111/stable-diffusion-webui, Accessed: 2025-05-31.

[51] *Verybadimagenegative - stable diffusion embedding*, https://huggingface.co/nolanaatama/embeddings/blob/main/verybadimagenegative_v1.3.pt, Accessed: 2025-05-31.

[52] *Bad-artist - stable diffusion embedding*, https://civitai.com/models/5224/bad-artist-negative-embedding, Accessed: 2025-05-31.

[53] *Bad-hands-5 - stable diffusion embedding*, https://civitai.com/models/116230/bad-hands-5, Accessed: 2025-05-31.

[54] L. Zhang, A. Rao, and M. Agrawala, *Adding conditional control to text-to-image diffusion models*, 2023. arXiv: 2302.05543 [cs.CV].

APPENDIX

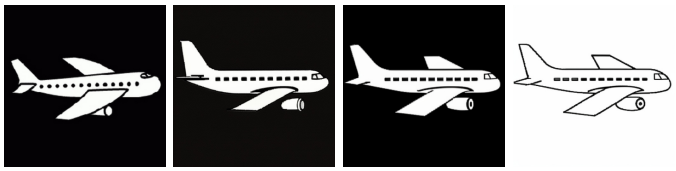

Figure 6. Concept *airplane* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 7. Concept *angry* (transparent). Left to right: original reference, text2img, refined black, refined white.

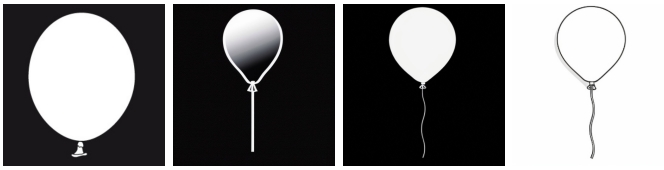

Figure 8. Concept *balloon* (transparent). Left to right: original reference, text2img, refined black, refined white.

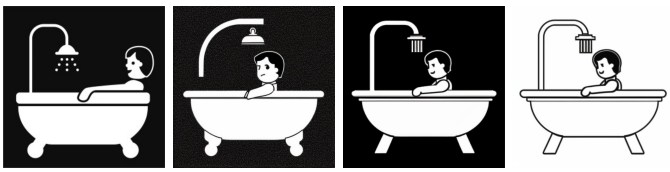

Figure 9. Concept *bathing* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 10. Concept *birthday* (transparent). Left to right: original reference, text2img, refined black, refined white.

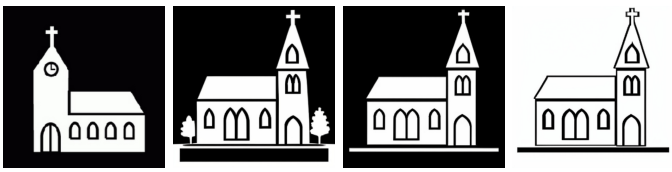

Figure 11. Concept *church* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 12. Concept *coffee* (translucent). Left to right: original reference, text2img, refined black, refined white.

Figure 13. Concept *cycling* (translucent). Left to right: original reference, text2img, refined black, refined white.

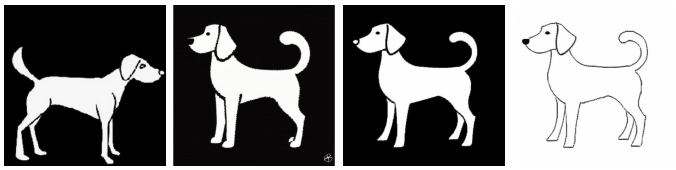

Figure 14. Concept *dog* (transparent). Left to right: original reference, text2img, refined black, refined white.

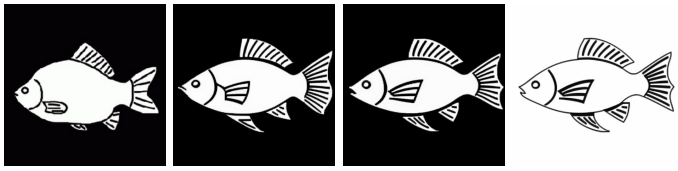

Figure 15. Concept *fish* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 16. Concept *headache* (translucent). Left to right: original reference, text2img, refined black, refined white.

Figure 17. Concept *hospital* (translucent). Left to right: original reference, text2img, refined black, refined white.

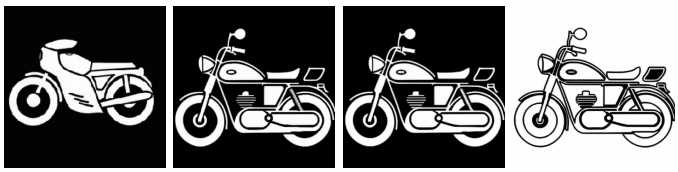

Figure 18. Concept *motorcycle* (transparent). Left to right: original reference, text2img, refined black, refined white.

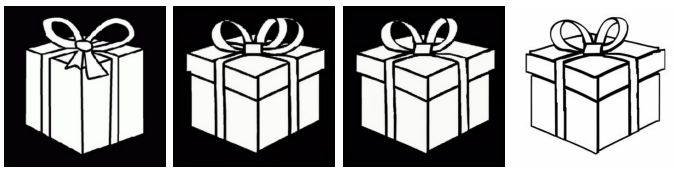

Figure 19. Concept *present* (transparent). Left to right: original reference, text2img, refined black, refined white.

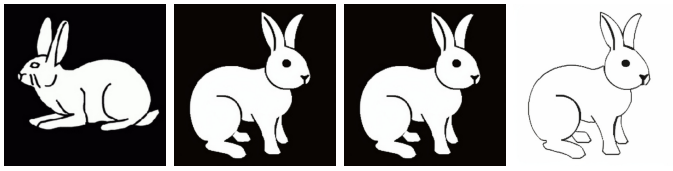

Figure 20. Concept *rabbit* (transparent). Left to right: original reference, text2img, refined black, refined white.

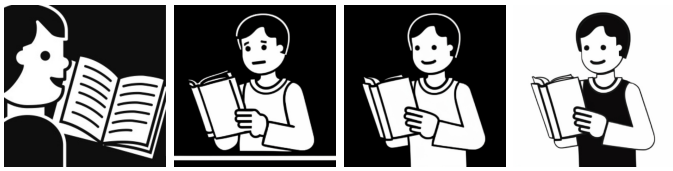

Figure 21. Concept *reading* (translucent). Left to right: original reference, text2img, refined black, refined white.

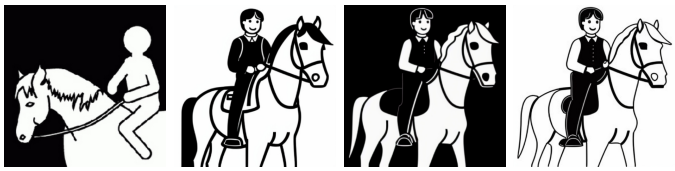

Figure 22. Concept *riding* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 23. Concept *sad* (transparent). Left to right: original reference, text2img, refined black, refined white.

Figure 24. Concept *thunderstorm* (translucent). Left to right: original reference, text2img, refined black, refined white.