Evaluating Diffusion-Based Image Generation for Easy Language Accessibility

Christoph J. Weber, Dominik Beyer, Sylvia Rothe

c.weber@hff-muc.de
chris-j-weber.github.io

Team & Research Focus

Christoph J. Weber is a PhD candidate in AI & Film at HFF & LMU Munich. With a background in Computational Linguistics and Computer Science (LMU), he explores human-centered generative AI systems that support authorship, control, and collaboration in creative media production.

Dominik Beyer developed this project as part of his Master’s thesis, investigating how AI can support inclusive media through simplified language and visual accessibility. He is starting a PhD at Inclumedia (ABM Medien) in the project “Easy Language and Visual Content: Using AI for Accessibility.”

Sylvia Rothe is head of the AI Lab at HFF Munich and leads research on AI in film, focusing on its practical implementation in everyday filmmaking workflows.

Easy Language

Generative AI models, particularly diffusion-based systems, enable the synthesis of highly detailed and photorealistic visuals from textual input. These models are capable of transforming abstract or narrative descriptions into coherent visual representations, opening new possibilities for creative expression and storytelling.

Complex Language

long sentences

hard to understand for people with:

- learning disabilities
- low literacy
- sensory impairments
- migration background

different tenses

Easy Language

Generative AI models, particularly diffusion-based systems, enable the synthesis of highly detailed and photorealistic visuals from textual input. These models are capable of transforming abstract or narrative descriptions into coherent visual representations, opening new possibilities for creative expression and storytelling.

AI tools can turn sentences like "a girl in a red dress" into real-looking pictures.
This technology is changing how people create visual stories.
A computer can do many things. It can make pictures.

logical structure

Plain Language

familiar words

clear sentence structure

active language

Easy Language

Generative AI models, particularly diffusion-based systems, enable the synthesis of highly detailed and photorealistic visuals from textual input. These models are capable of transforming abstract or narrative descriptions into coherent visual representations, opening new possibilities for creative expression and storytelling.

AI tools can turn sentences like "a girl in a red dress" into real-looking pictures.
This technology is changing how people create visual stories.
A computer can do many things. It can make pictures.

You write a sentence.
For example: “A girl in a red dress.”
Then the computer shows the picture.

Easy Language

DIN SPEC 33429 (Germany)
Inclusive Communication Hub, Easy Read UK
(not formally reviewed => not “Easy Read”)

Easy Language + Images

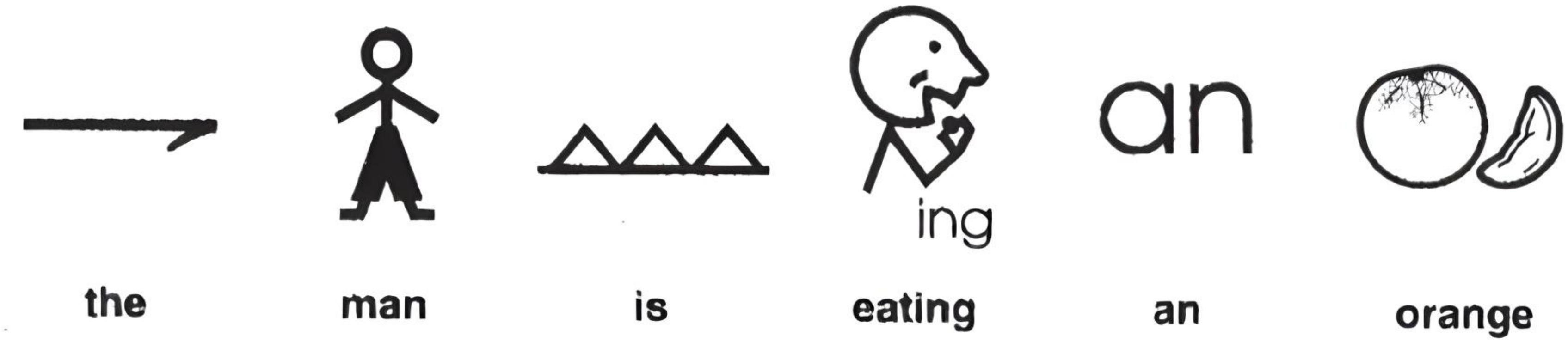


Illustration of opaque (“the”, “is”), translucent (“eating”), and transparent (“man”, “orange”) symbols used to support sentence comprehension [1].

[1] 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. 466474, 2007. DOI: 10.1111/j.1468-3148.2006.00334.x.

Problem and Motivation

⚠️ **Problem:**

Manual symbol design = slow & expensive

❓ **Research Questions:**

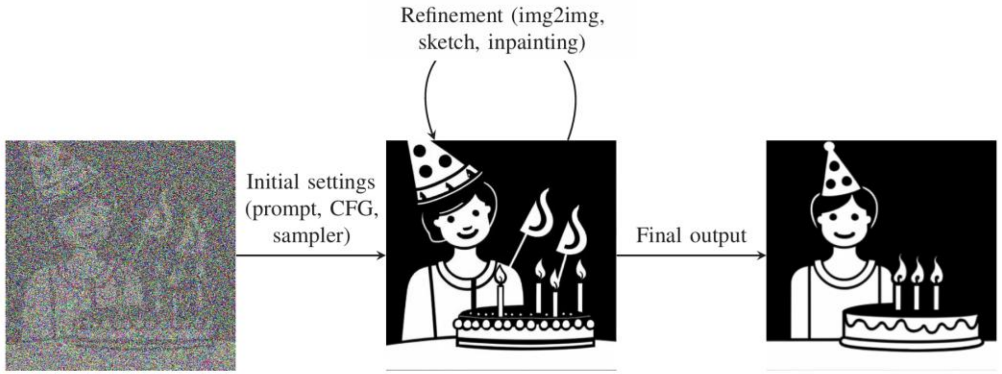
- Can a diffusion-based model reproduce a consistent Easy Language visual style?
- Are the generated images semantically clear for the target audience?

Dataset & Fine-Tuning Setup

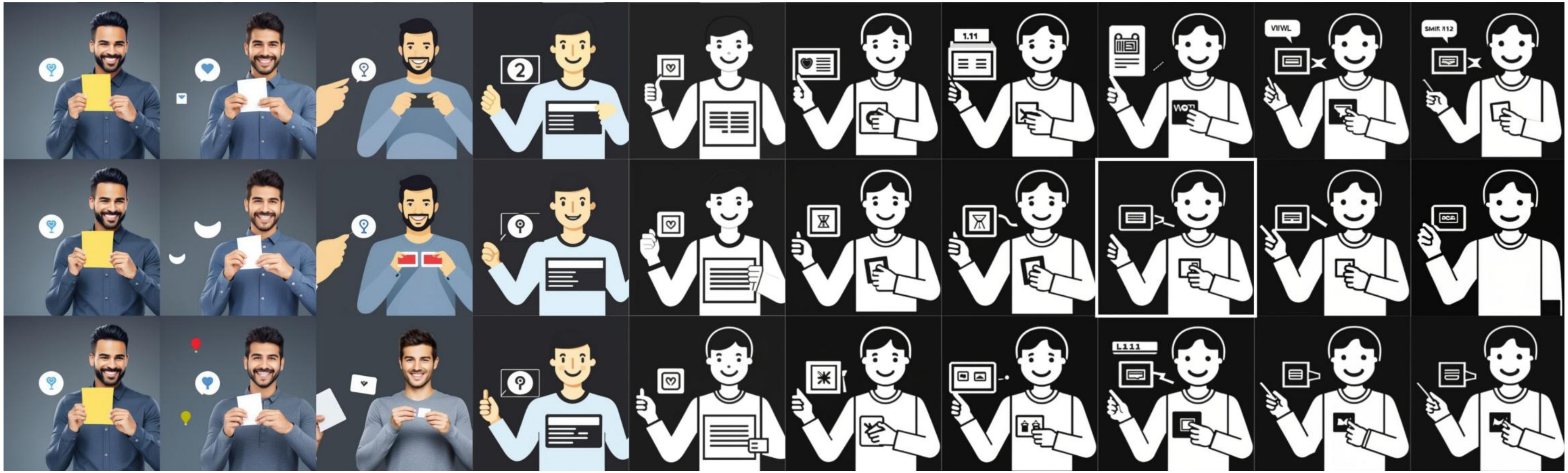
🗂️ Source:
99 symbols from Picto-Selector (Pictogenda subset)
→ Black-and-white, minimalist, style-consistent

🏷️ Prompting:
Custom captions incl. style token and layout tags
"illustration in the style of pl41nl4ng, {detailed description of the image}, {additional tags}"

🧠 Model Setup:
Base: Stable Diffusion v1.5
Fine-tuned Stable Diffusion Model using LoRA
30 epochs · 29.7k steps · 10× image reuse



Training Results



🎯 Model selection:

- Epoch 19, strength 0.8
- Best stylistic fidelity
- Clean outlines, correct layout, no artifacts

Study 1: Visual Distinctiveness

🎯 Goal:

Can users distinguish AI-generated images from original pictograms?

🧪 Design:

- 20 image pairs (AI vs. original)
- Binary classification & preference rating
- Based on transparent and translucent concepts

👥 Participants:

15 participants (mostly students, 6 with experience in AI image generation)

Study 1: Visual Distinctiveness

Distinctiveness Accuracy: 47.7% (→ at chance level)

Image preference:

- Original images: 10.3%
- AI-generated images: 40.0%
- No preference (ties): 49.7%

Reported judgment criteria:

Line sharpness, face details, object proportions

Study 2: Semantic Clarity

🎯 Goal:

Can users correctly interpret the meaning of AI-generated symbols?

🧪 Design:

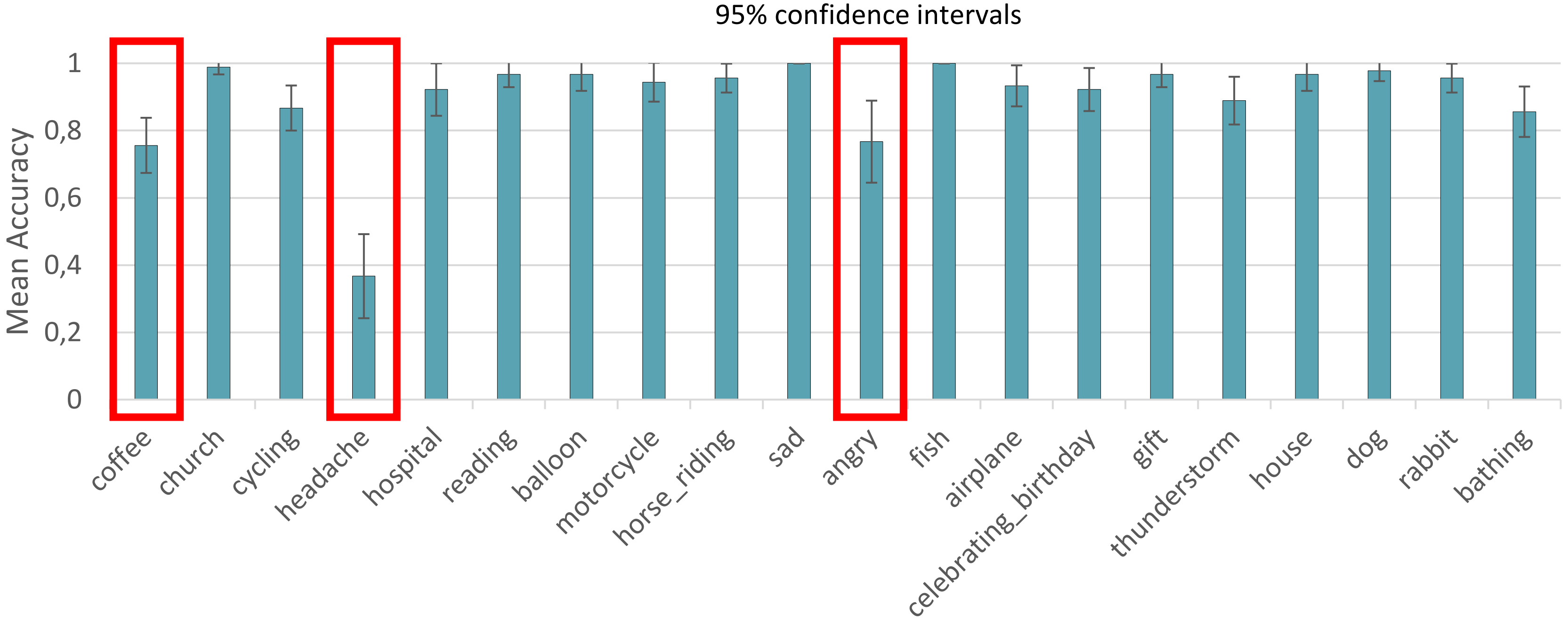
- 20 images (10 transparent, 10 translucent)
- Open-ended text input (max 100 characters)
- Answers scored: 1 = correct, 0.5 = partially correct, 0 = incorrect

👥 Participants:

42 total, including Easy Language testers and users with learning difficulties

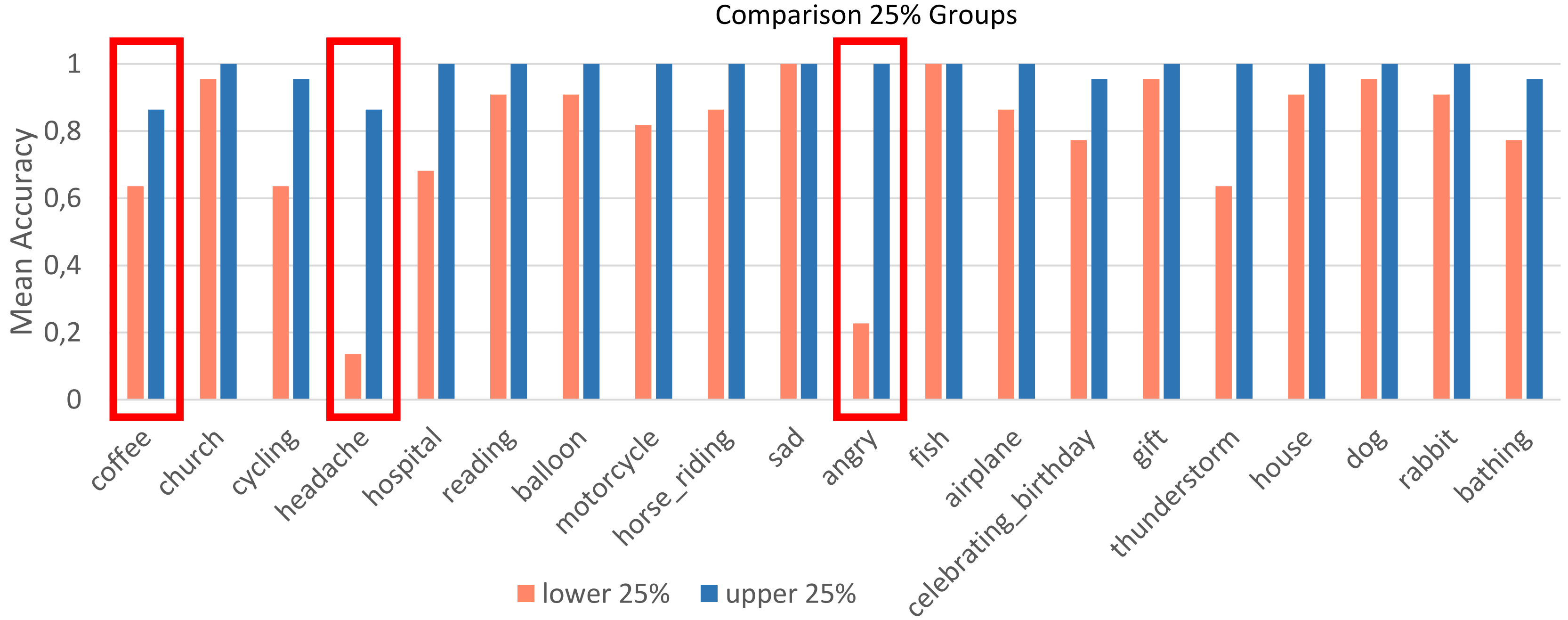
Study 2: Semantic Clarity

- ✅ Overall accuracy: 89.8%
- 🔹 Transparent concepts: 94.6%
- 🔹 Translucent concepts: 85.1%
- 🔹 Significant differences between concepts (Friedman test: $p < 0.001$)



Study 2: Semantic Clarity

- Lower 50% performing group: acc = 83.0%
- Lower 25% performing group: acc = 77.7%



Other Qualitative Insights

💬 User Feedback:

- Some AI-generated images felt “*too childlike*”
- Some questioned *whether certain concepts needed images at all*

❗ Content-related challenges:

- Abstract or emotional meanings (e.g., headache, angry)
- Especially problematic for lower-performing users

🔍 Takeaway:

Recognition depends on both symbol design and user background

Discussion & Limitations

✅ Successes:

- Style fidelity achieved with a small dataset
- High semantic clarity across most concepts
- AI outperformed originals in preference task

⚠️ Limitations:

- Abstract/emotional concepts remain difficult (e.g., headache, angry)
- User study did not include only Easy Language target users
- Minimalist style only → unclear generalizability to other styles

🔁 Human-in-the-loop remains essential for quality control and iteration

Conclusion & Future Work

Conclusion

✅ Diffusion models can support Easy Language

stylistically consistent and semantically clear visuals

🧠 Fine-tuning with LoRA

enables strong results even with small datasets

📊 User studies show

- AI-generated images were not distinguishable from human-made
- Human-made images were not preferred over AI-generated
- AI-generated images are generally easy to understand
- Abstract/emotional concepts remain a challenge

Future work

🧪 Focused and more detailed evaluation

with specific subgroups of the Easy Language audience

🧠 Improving abstract/emotional concepts

Better prompting, fine-tuning, and model steering for nuance

⚙️ Controllable models

Leverage new AI models for better precision

🛠️ Tool development

Interfaces for translators: “highlight text → generate image”

HOCHSCHULE FÜR FERNSEHEN UND FILM MÜNCHEN

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Questions ?

c.weber@hff-muc.de

IARIA