

Image Generation for Work Instructions

Flexibele Werkinstructies

Swarupa Hardikar

Is it possible to generate diagrammatical representations of work instructions for a more comprehensive visual aid? The explorative study tests this with four state-of-the-art image generation AI tools.

Experimental Setup

- An example work instruction for building a blue wooden table, consisting of four steps
- The four models, prompted to generate an IKEA-style representation of work instructions
- 2 modes of prompting: A single prompt for all 4 steps to create 1 image vs. A different prompt and image for each step
 - The first mode was not utilized due to the introduction of significant hallucinations in the image, rendering it completely unusable.

Prompts

- To ensure ideal implementation, I prompted ChatGPT to rewrite the work instructions to suit the input of an image generator model
- Prompt template: "Illustrate a step-by-step IKEA-style instruction for assembling a blue wooden table. The style should be clear and minimalistic, without any text or labels. Each image should depict only the relevant components and tools in use. [INSERT STEP]"
- This template is used for every generation, followed by the instruction.



Example Single-prompt output

PROMPTS

"A single wooden plank (20x50cm) lies flat on a workbench. A saw is actively cutting it in half lengthwise. Two shorter planks (20x25cm) are placed beside the plank, showing the cutting process"

"Four wooden planks are aligned side by side on a workbench. A hand is using a hammer to drive a nail into one of the planks, connecting them together to form a tabletop."

"A wooden tabletop is flipped upside down. Four short wooden legs (20x25cm) are positioned at each corner. A hammer is being used to drive nails into one of the legs, securing it to the tabletop."

"A completed wooden table stands upright. A person's hands are pressing down on it to check for stability. A hammer and extra nails lie nearby, indicating final adjustments if needed."

Findings

Stylistic differences: For most models, generating step-by-step multi-prompts differences causes stylistic differences across the illustrations.

Ignoring style instructions: Both Firefly and Copilot ignore the instructions to create an IKEA-style, minimalistic illustration, instead generating realistic images.

Hallucinations: A lot of visual "hallucinations" can be observed for most of the models as well, such as misshapen table legs, incorrectly generated text, missing tools, etc.

Best performance: Out of the 4 models, the best-performing one by far is **ChatGPT**, which adheres to most of the instructions, recreates the IKEA style and shows minimal hallucinations.

Worst performance: Notably, the worst-performing in this experiment was **Midjourney**, which does not adhere to multiple instructions (such as not generating text or applying a minimalistic style) and produces multiple hallucinations.

Conclusion & Future Directions

AI image generation tools show promise in generating work instructions, but the hallucinations as well as lack of clarity may cause safety issues.

On the other hand, the ChatGPT generation, which is the most recent (April 2025) already shows huge improvements, indicating that the accuracy of image generators may improve in the near future.

The experiment used IKEA-style instructions as a guide; however, other illustration styles can also be tested.

MODELS

Adobe Firefly



Midjourney



MS Copilot



ChatGPT

