

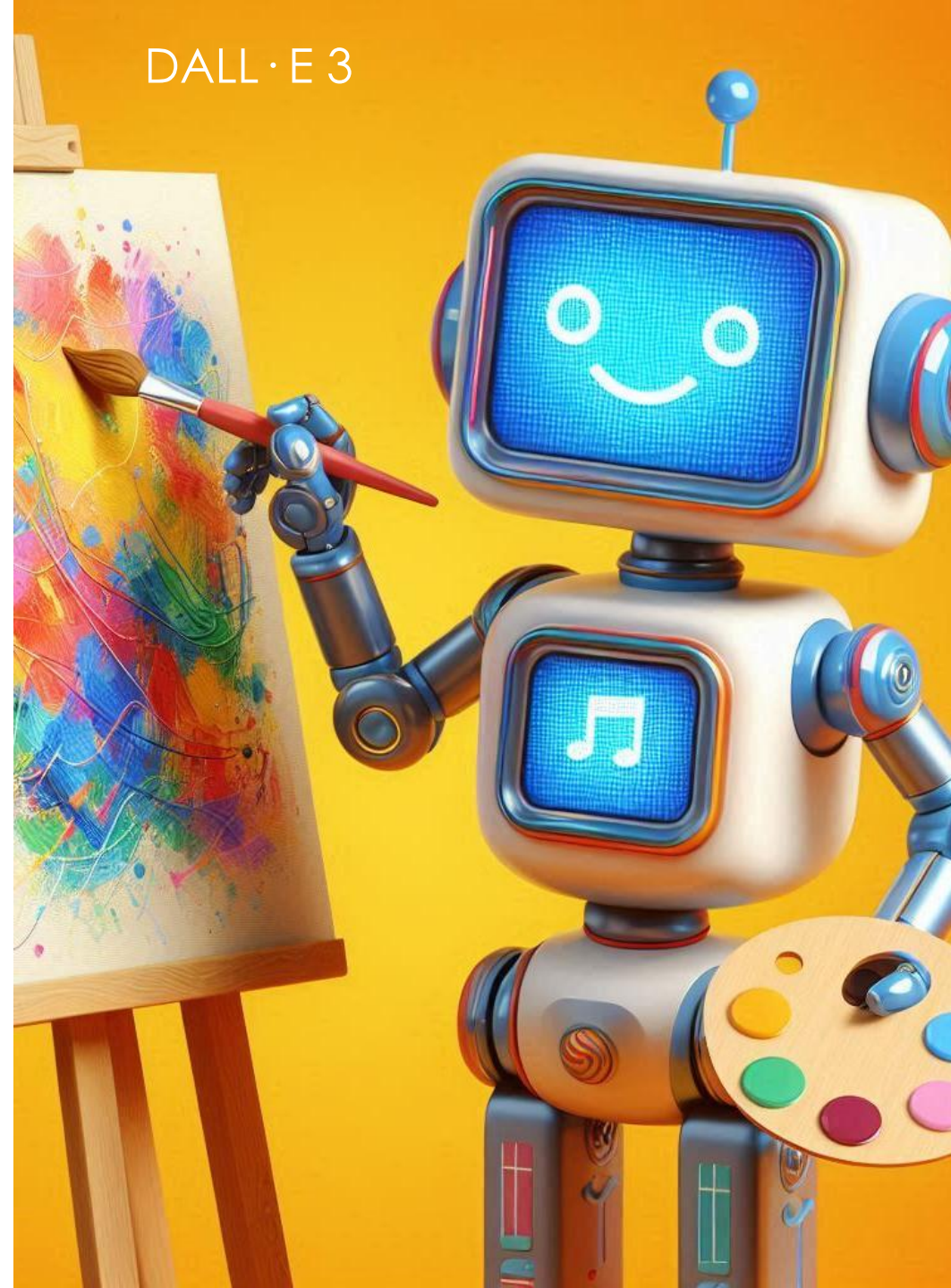


GENERATIVE
AI APPLICATIONS

Generative AI applications

- ▶ Generative AI (GenAI, GAI) applications are applications that generate images, text, audio, or videos based on a given prompt.
- ▶ The ability of generative AI applications to generate new content is based on the training material. Therefore, all the content generated by generative AI applications reflects the training material used.

DALL·E 3

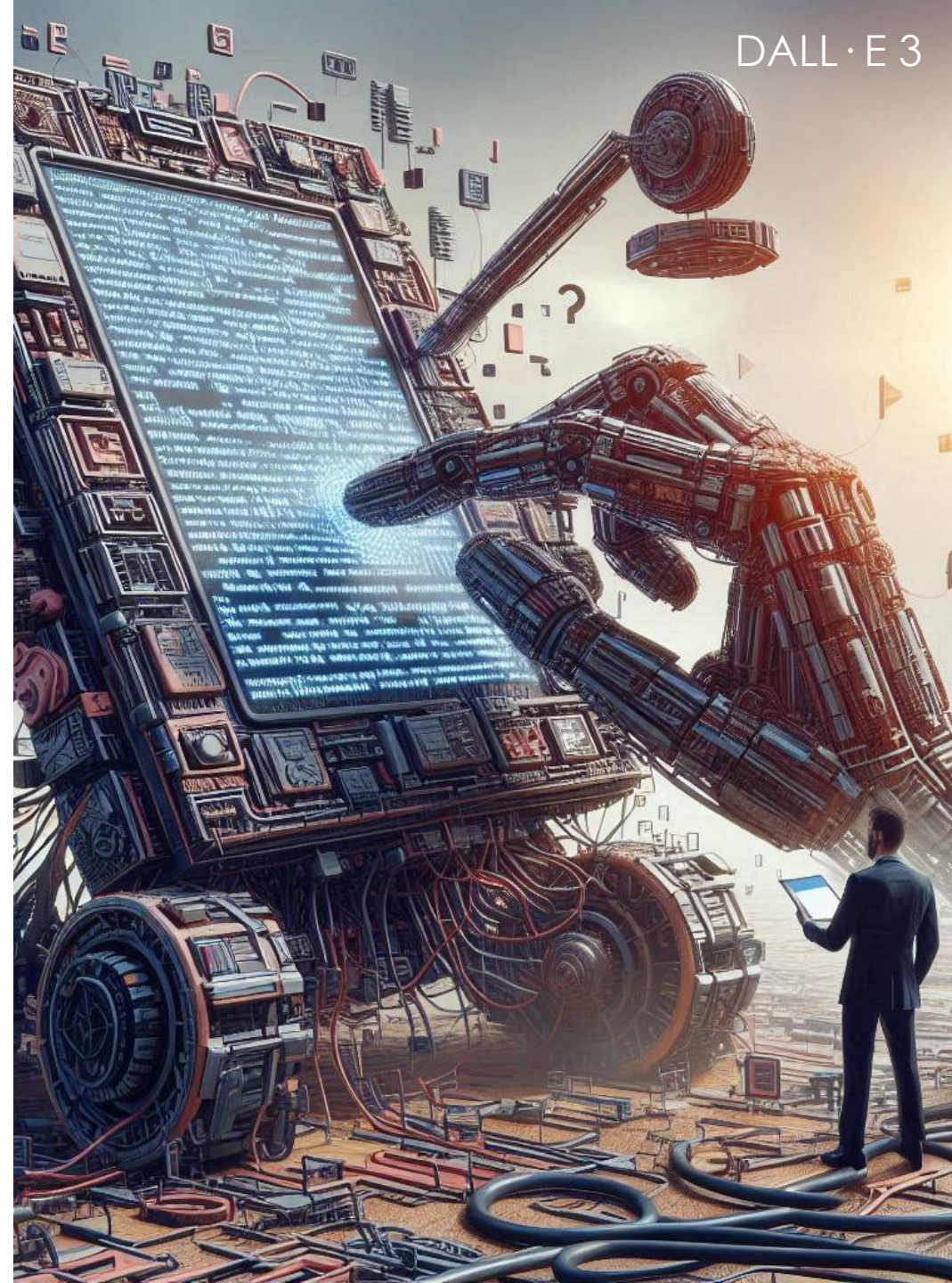




CONVERSATIONAL
AI APPLICATIONS

Large language models (LLM) – the foundation for conversational AI applications

- **Conversational AI applications are based on large language models (LLM).**
- **Large language models are neural networks** that have been trained on a large amount of training material that contains natural language. The largest language models have trillions of parameters.
- In addition to training materials, large language models are trained with human feedback.
- Language models are language-specific. The largest language models are based on the English language.
- The language model does not even try to understand the text it reads, but only tries to continue it in the most probable direction determined by the language model.

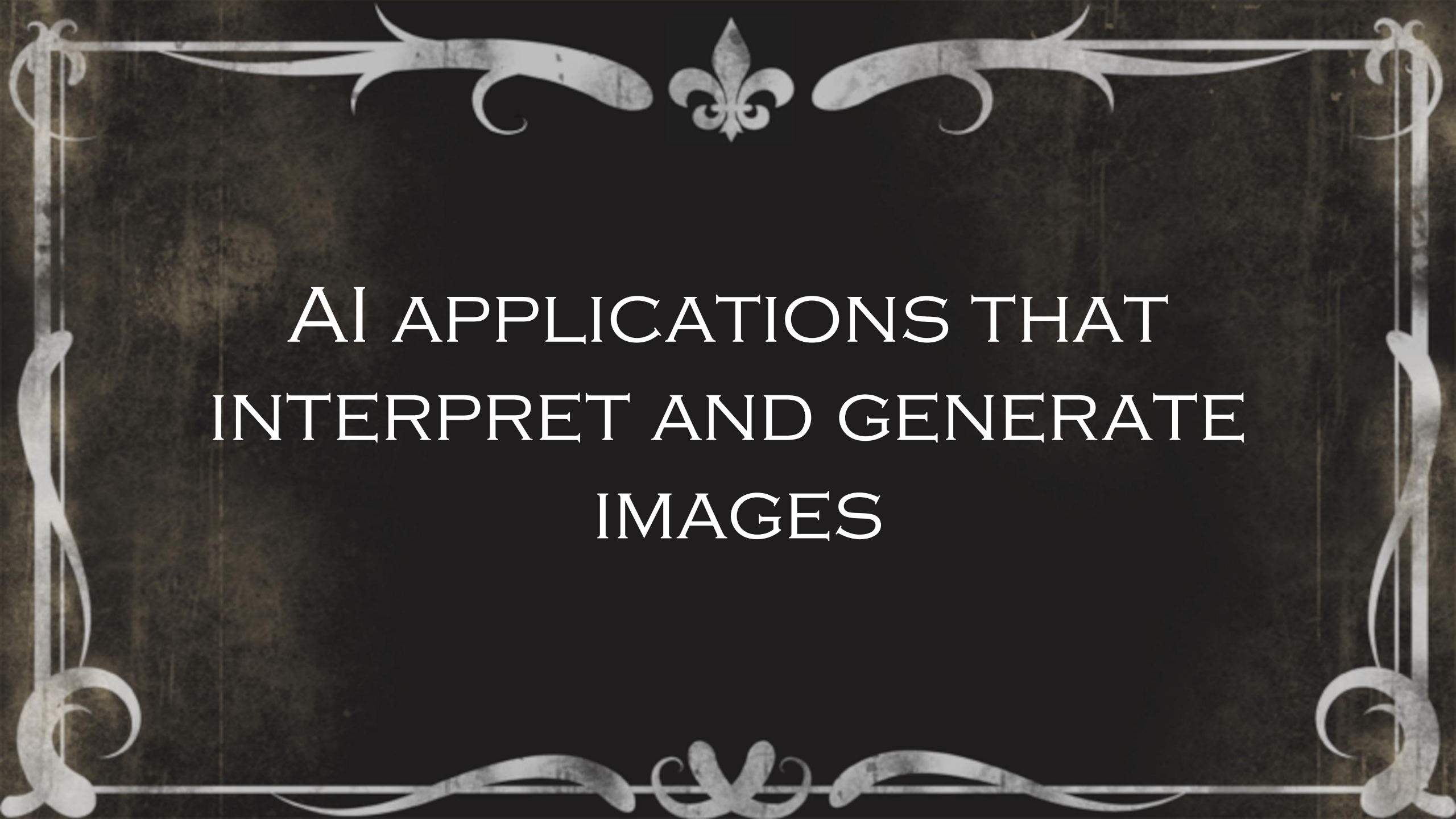


Conversational AI applications

In addition to large language models, conversational AI applications have other functions for different tasks. Some of the functions are neural network-based and some are implemented with rule-based programming. These functions are usually poorly documented, but we can assume that the applications will include separate functions such as:

- **Prompt preprocessing:** To identify specific factors of the prompts, they are analysed with a rule-based approach or neural networks that are trained for preprocessing. One example of this is the analysis to identify the user's goal based on the prompt (*intent analysis*).
- **Solving mathematical expressions:** Solving mathematical expressions is done significantly more efficiently and reliably with a separate rule-based solution than with a neural network.
- **Text recognition (OCR):** Rule-based text recognition is also used to support neural network technology.
- **Grammar, language translations:** Inaccurate results suggest that grammar checking and language translations are mainly done on a neural network basis.
- **Moderation and censorship,** likely at many levels of the application.
- **Real-time information retrieval:** Retrieving information directly from the internet or a database to complement the given prompt.

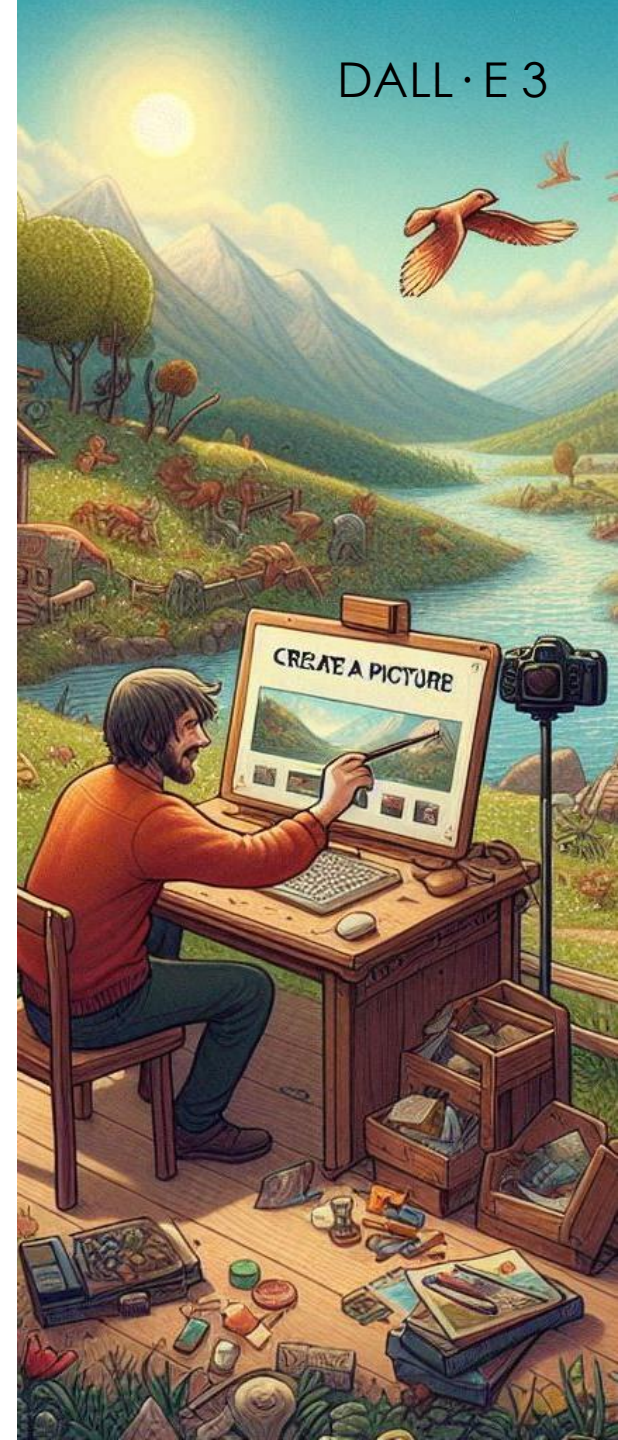




AI APPLICATIONS THAT
INTERPRET AND GENERATE
IMAGES

About image processing in generative AI applications

- Generative AI applications already interpret and generate images quite well.
- Image processing is also based on neural network technology, but what is special about image processing are the highly developed mathematical methods that refine training material for neural networks.
- A prerequisite for interpreting and creating images is that the images used as training material are accompanied by textual information about the content of the image. Obtaining correct and sufficient text descriptions for images used as training material is a significant challenge for the development of image processing.
- Image creation can be improved with human feedback, for example, so that the model produces a series of images from the same input, and the person chooses the image from the series that they think best meets certain pre-set criteria.



Example of image interpretation

The AI told us the following about this real photo:

"The photo appears to be a black and white photograph with a red car in the middle. The car is a vintage Fiat 500, and it has a bouquet on its hood. The couple stands in front of the car and kisses. In the background, a cobblestone street and buildings are visible. Based on the photograph, it seems to have been taken in Europe, because the license plate and architecture suggest it."



Example of image creation and interpretation

Artificial intelligence created an **image of a chorus of butchers**. Below is the AI's interpretation of the image it created.

"In the picture, there appears to be a group of people in white lab coats and hats holding various tools in their hands, such as saws, hammers and axes. This description may refer to some theme or event where lab workers' outfits are combined with unusual objects, such as tools. It can be part of an art project, a theatre performance or other creative expression."



Integrity errors in image creation

AI does not "understand" what it is drawing, but develops and continues the pixel mass of the image statistically based on the given prompt and the training material used. This process is steered in the right direction by the intersection of the prompt given to the AI and the image's text description in the training material.

When this picture of a choir of psychotherapists was created by AI, it did not understand that it was drawing humans, and, for example, that humans have two hands and five fingers in each hand.

The image shows errors which are a result of this challenge.



Integrity errors in image creation

In this AI-generated image, the groom is threading the ring onto the bride's finger.

This example also shows a typical integrity error. The error occurs when an AI application tries to "continue" the image according to the content defined by the prompt, without knowing that humans have two hands.

However, the error blends smoothly into the image for the same reason, i.e., because the application tries to create a typical continuum in all parts of the image.

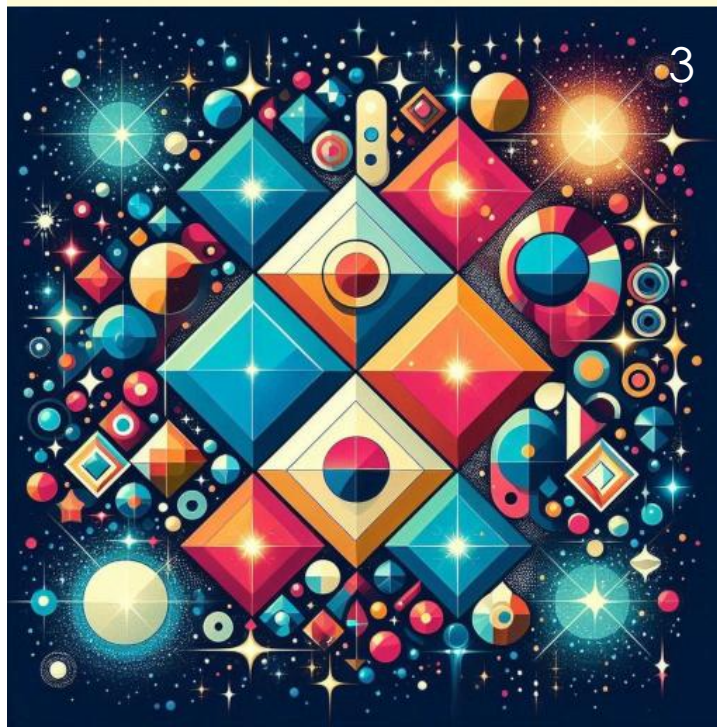


Integrity errors in image creation

Artificial intelligence creates images as a continuum of the pixel mass, not, for example, by supplementing an image one object at a time. This mechanism does not support number management.

As an example of this challenge, is the following images, on which the AI was asked to draw:

1. three cars, three bicycles and three pedestrians.
2. three apples, four coins and two pens.
3. four squares, four circles and four stars
4. five cats, five dogs and five rabbits.



Content biases in the image creation

- ▶ The images generated by AI are reflections of the mass of images used as training material. For this image mass to well represent the real world, it should consist of completely random shots of the real world. Of course, this is not the case, since the images have been created to serve different purposes and have been selected for publication on various grounds.
- ▶ For AI to be able to make full use of the mass of images, there should also be a comprehensive and objective text description of the content of the images. This is rarely realised, as a text description usually serves the needs of the publisher at the time of publication.
- ▶ These factors cause biases in the images generated by artificial intelligence in relation to the real world.



Content biases in image creation

Bias is caused, for example, by the fact that **images that present the most relevant, well-known, representative or interesting aspects of the subject are selected for publication.**

However, in the training material, these images represent the entire world for AI.

When AI creates images of Helsinki, for example, it persistently tries to capture Helsinki Cathedral or even several in the image.

The prompts for these images requested photos of different districts of Helsinki in different years. None of the prompts mentioned the Cathedral.



Content biases in image creation

AI drew a picture of "**Firefighters spending summer vacation on the beach**".

If the description of a picture in the training material mentions a firefighter, the person is likely to be in firefighter gear and performing firefighter duties. But, if a firefighter is, for example, in a vacation photo, it may not be mentioned that he or she is a firefighter.

The absurdity of the image is a consequence of this bias in the training material regarding the concept of "firefighter".

The same bias can occur in other similar concepts, such as different professions, ethnic groups, etc.





GUIDELINES FOR USING
GENERATIVE
AI APPLICATIONS

On the reliability of responses from generative AI applications

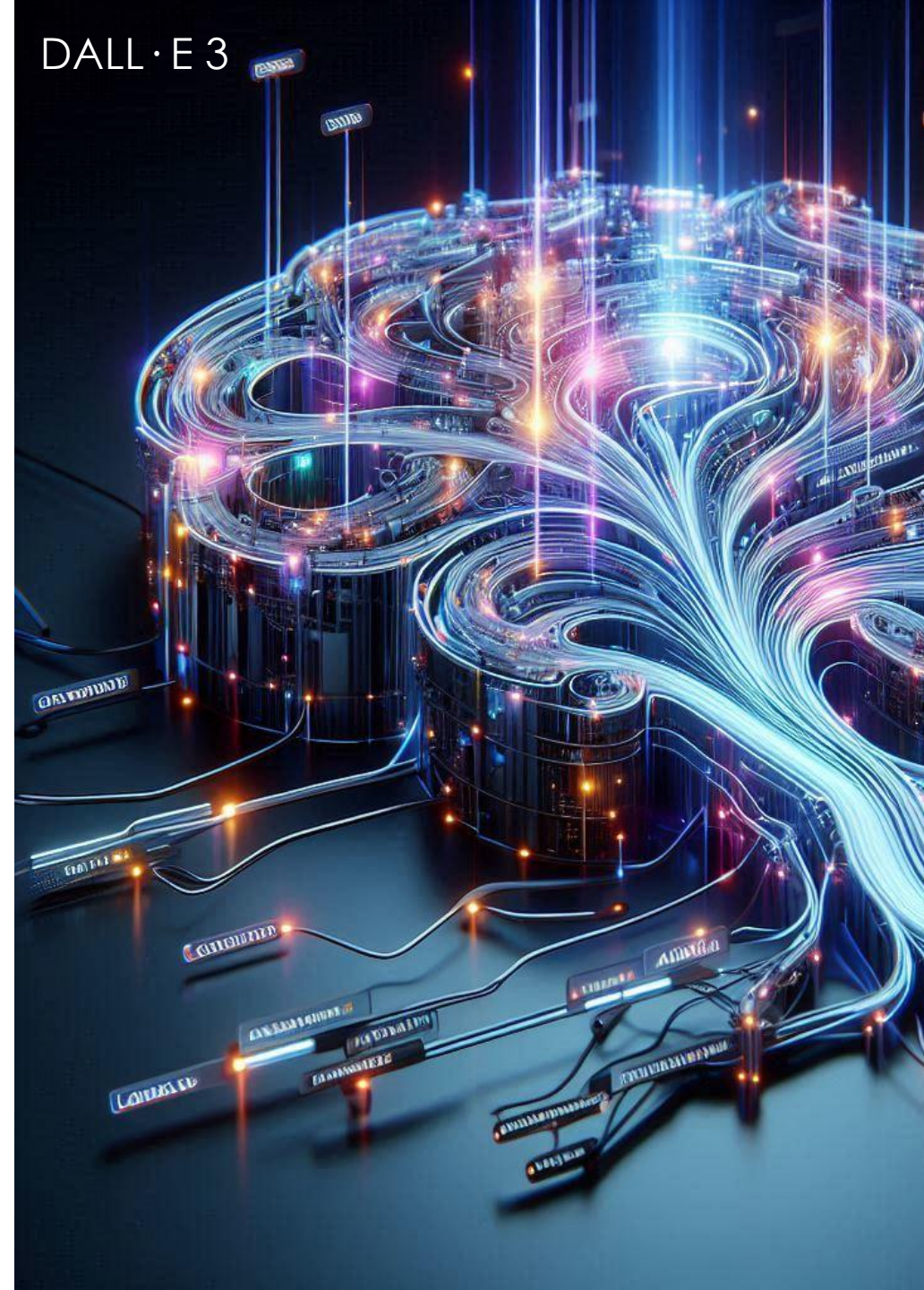
The text generated by language models is usually quite convincing. This easily creates the illusion that the text is an accomplishment of an intelligent actor. However, the text is a statistical prediction, which is mathematically calculated based on the training material - not a result of any reasoning.

So don't let persuasive language fool you; the factual content may still be incorrect or biased.

Also, don't be discouraged by incorrect information; another assignment can yield valuable results.

Actually, generative AI should not be viewed as a tool that gives right or wrong answers, but as a tool that gives the best possible approximation or prediction based on the training material and prompt that is offered to it.

However, the user of generative AI must always be able to assess the quality and reliability of the results themselves.



Guidelines for using generative AI applications

- ▶ AI gives a more accurate result, the better the gripping surface the prompt gives to the training material. This is affected by, among other things, the content, structure, detail, word choice and style of the prompt text.
- ▶ The biases in AI outputs, which were easy to demonstrate with image examples, also apply to textual outputs. Biases and integrity errors cannot be completely avoided, even with a good prompt.
- ▶ AI applications may also retrieve information directly from the internet or other sources outside the neural network. However, the information retrieved is not associated with the response, but as part of the prompt. In other words, it becomes part of the context provided to the neural network, not part of the answer. Therefore, there is no guarantee that the information retrieved from the internet or other external sources ends up unchanged or at all in the answer given by the application. For the same reason, the internet references should not be presented as source information for the answer.



