

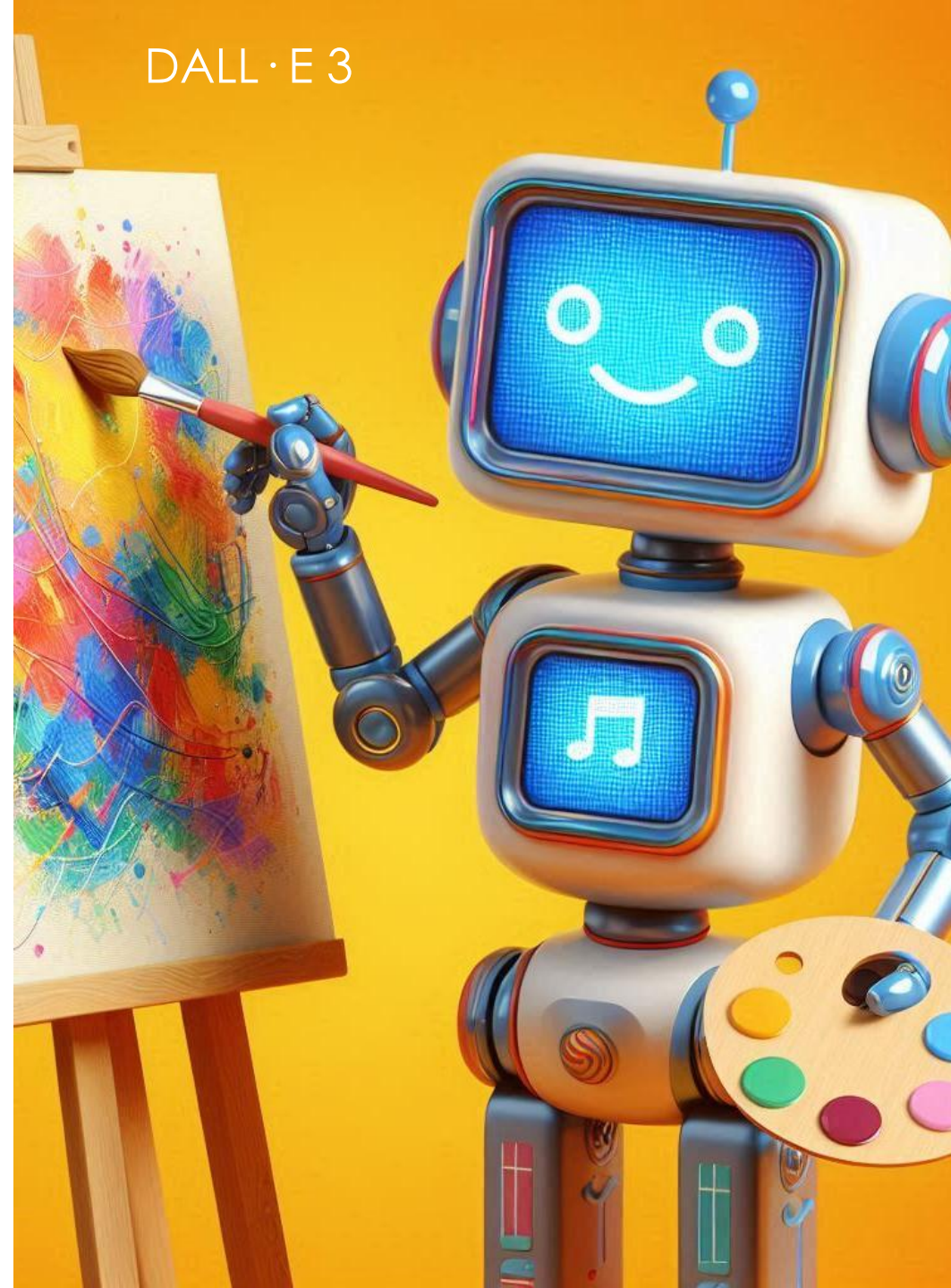


GENERATIVE
AI APPLICATIONS

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

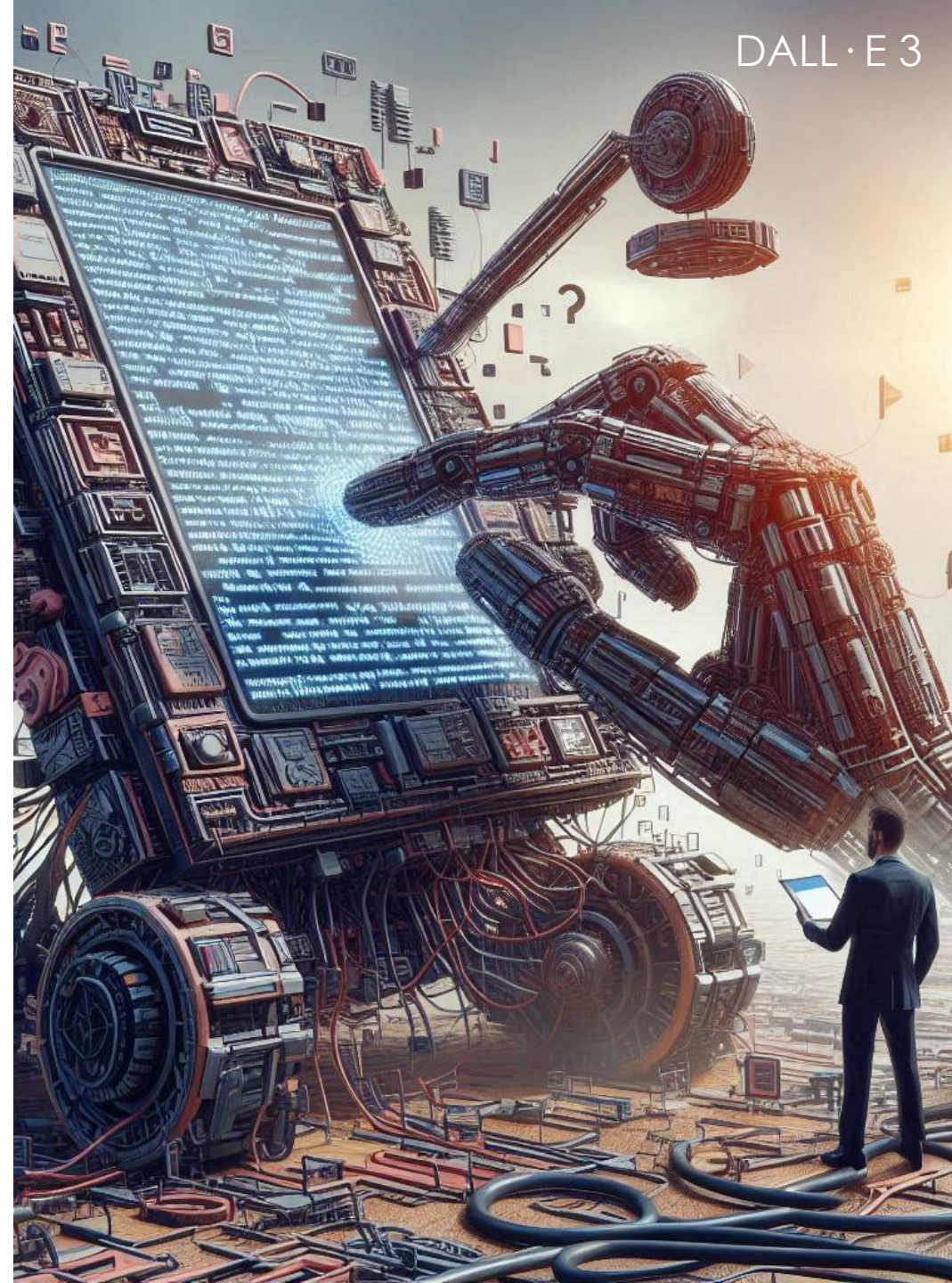




CONVERSATIONAL
AI APPLICATIONS

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 language model does not contain grammar or vocabulary** that humans can interpret, but the structure of language and the semantics of words are implicit statistical models built from teaching material.
- **The language model does not understand the text it reads**, but continues it in the statistically most probable way. Even the model's apparent reasoning is only a simulation of the textual structures that contain the reasoning in the training material.



Multilingual Large Language Models

- Many large language models are multilingual, i.e. their training material includes sources written in several languages. Different languages and structures common to several languages affect the values of the parameters of the neural network and the performance of the model more strongly, the more often they appear in the training material.
- In a way, languages disappear in neural network training because the parameters of the network do not contain information on how different languages and structures have affected their values. The same parameters serve several languages at the same time. Therefore, in multilingual language models, minority languages can benefit from the structures of majority languages.
- Translation between different languages takes place in multilingual language models without a separate process within the model through the grammar and vocabulary implicitly stored in the parameters of the neural network. In monolingual language models, translation from one language to another is done outside the language model.



Conversational AI applications

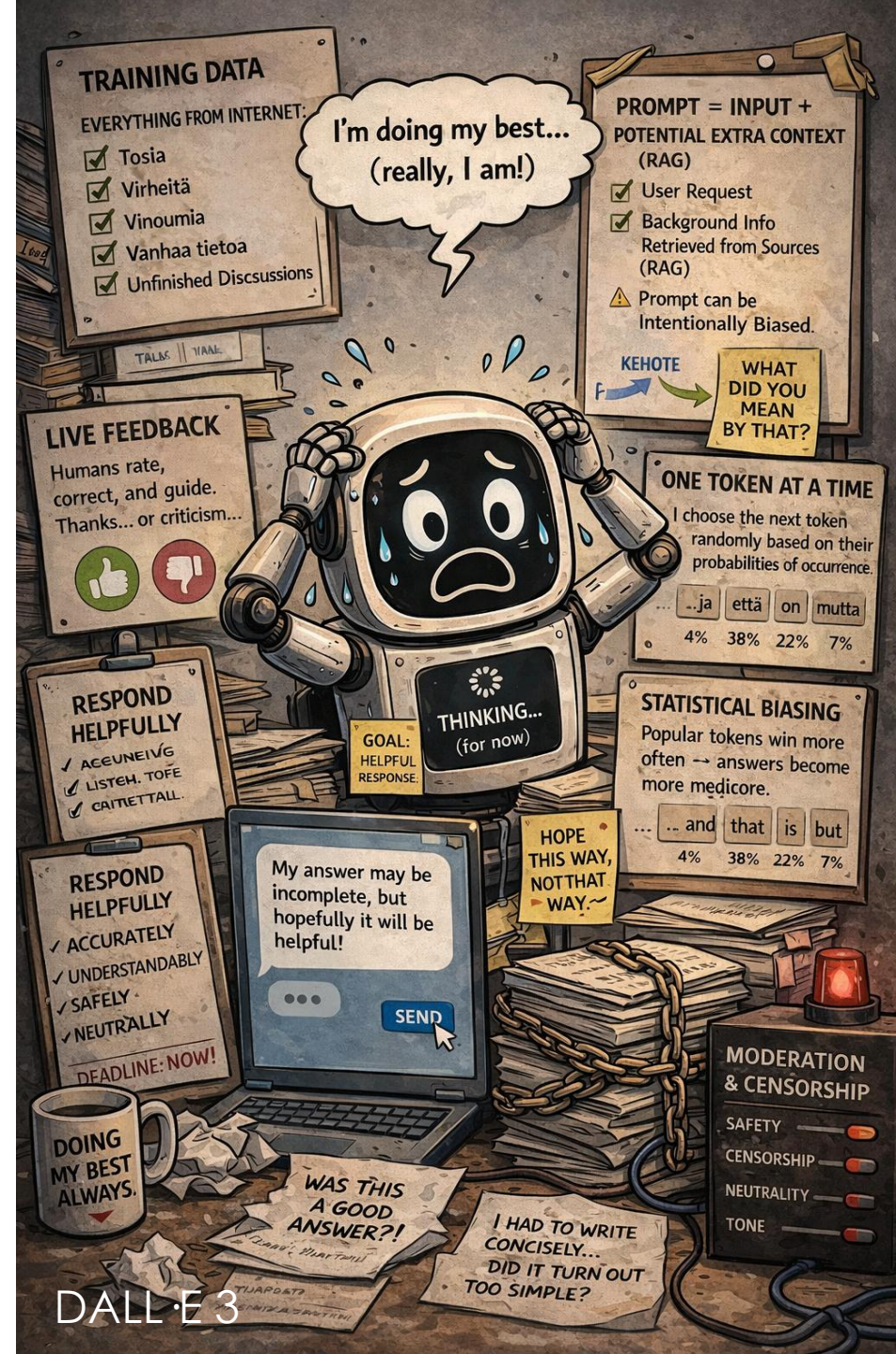
In addition to language models, conversational AI applications consist of several other components: neural networks, rule-based systems, and external tools.

- **An orchestration layer** decides which tools and functions to invoke in each situation.
- **Memory and context management** are essential to the system as a whole.
- **User identification** retrieves the settings, permissions, or profile associated with a user before processing begins.
- **Prompt preprocessing** performs various tasks, such as the analysis of the user's intent.
- **E.g. mathematical problems and program code syntax checks** are handled more reliably by rule-based systems than by neural networks.
- **Retrieval-Augmented Generation (RAG)**: The application can retrieve information in real time from the internet or a database. However, the retrieved information is not inserted directly into the response, but injected into the prompt, allowing the language model to draw on it when formulating its answer.
- **Content moderation and filtering** often occur in multiple stages.



Factors affecting the responses of a conversational AI app

- The language model is trained with **training materials** and **often with human feedback** (RLHF, Reinforcement Learning from Human Feedback).
- **The language model's response is based on the prompt.** The AI-application can add information from an external source (RAG, Retrieval-Augmented Generation), the model developer's own instructions, or even deliberately skewed information.
- The application builds its answer one token at a time by **randomly selecting the following tokens from the probability distribution of their occurrences.** However, **the mathematics of language models favours relatively more tokens with a high probability of occurrence.** This mediocres the answers and reinforces biases.
- The app's responses can be influenced by different levels of **moderation and censorship, as well as other separate components.**



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.

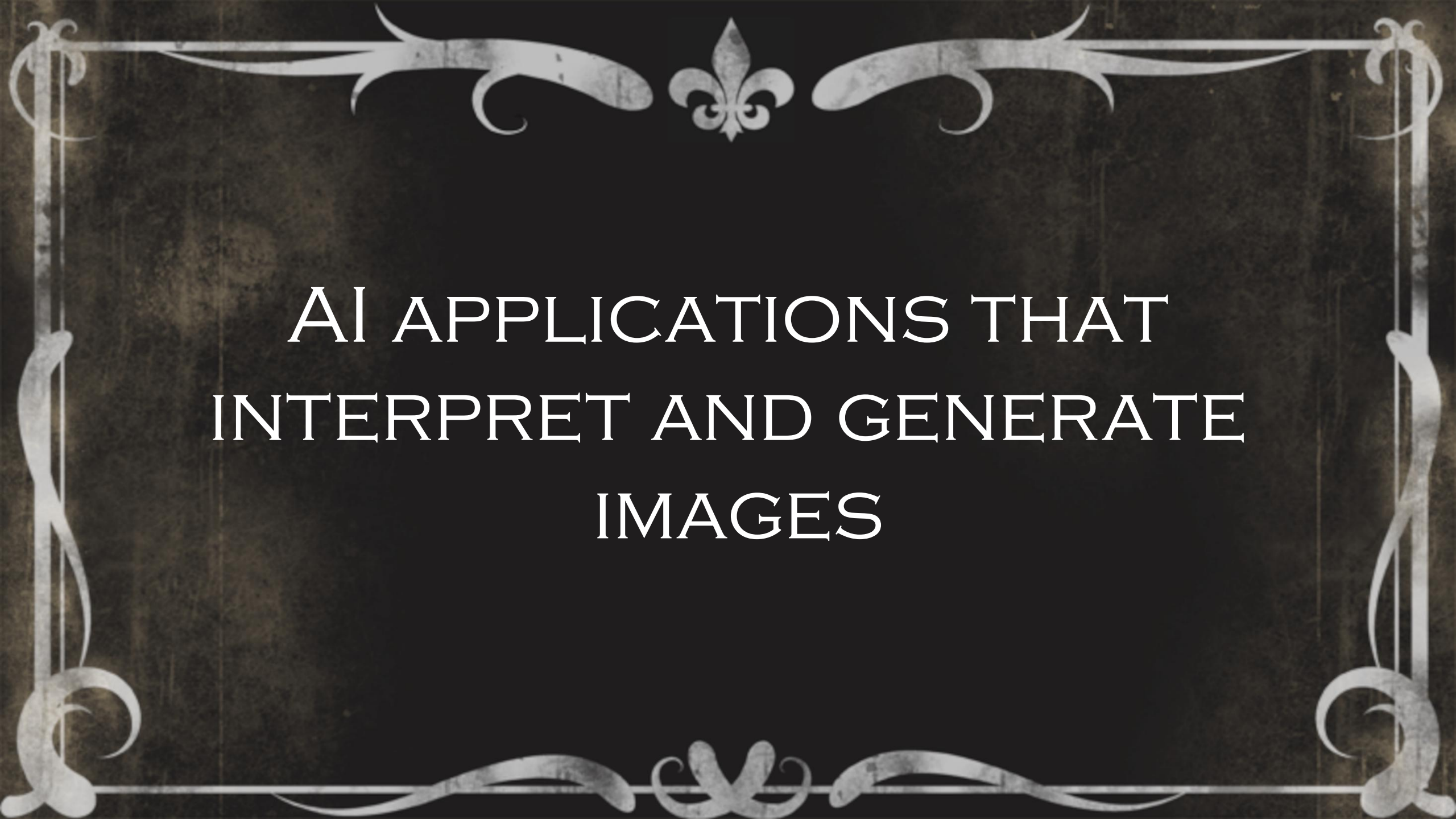
So don't let persuasive language fool you; the factual content may still be incorrect or biased.

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.





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.



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 interpretation

May Day in the Snowfall (Claude)

- ▶ The picture is about a Finnish May Day in a snowfall.
- ▶ A person sells or holds balloons.
- ▶ There is another person walking in the background, possibly a customer.
- ▶ The location is probably Helsinki – the urban environment and an old vehicle can be seen in the background. It could be, for example, the Market Square or some other central place.
- ▶ Date: May 1, May Day, but the year is a guess. The black and white and the style of the cars refer to the 1970s.



*Selling May Day balloons in 1967 in Helsinki.
Picture: [Erkki Laitila](#)*

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.

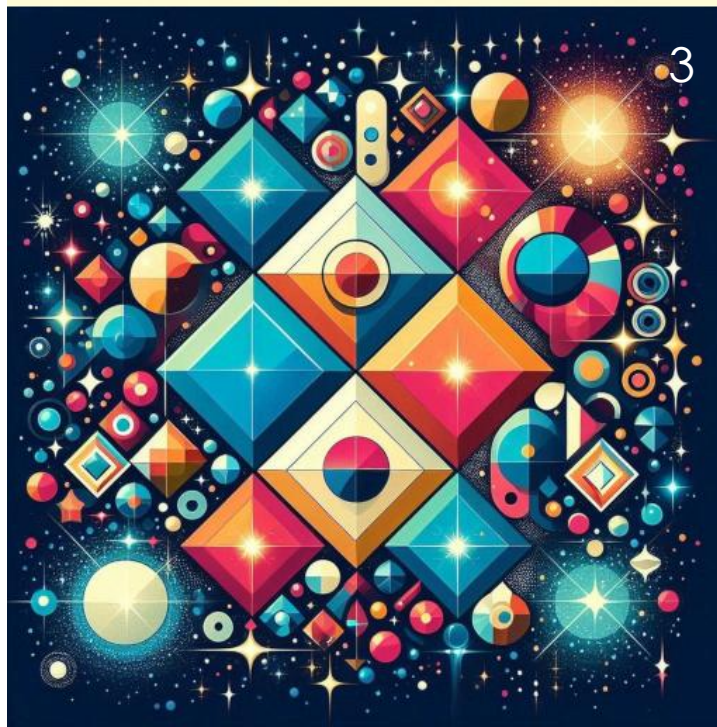


Integrity errors in image creation

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.

Therefore, in the worldview of AI, a firefighter may be born in full firefighter gear.

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



Content biases in image creation

Biases similar to the summer holiday image of firefighters can also occur in other contexts, such as professions, cultures, and ethnic groups. An example of this is the image drawn by artificial intelligence: "**Doctors spending their summer holidays on the beach**". This image shows both occupational and gender bias.

Biases that are easy to demonstrate with images also apply to text, voice, and other generative AI outputs.

Biases cannot be completely avoided even with a good prompt, and they are not always as clearly recognisable as in the examples.





GUIDELINES FOR USING
GENERATIVE
AI APPLICATIONS

