

How generative AI makes your life as a data engineer easier

Alexander Klein

Data Saturday #50



Sie

male mir ein Bild zum 50. Geburtstag des Data Saturday

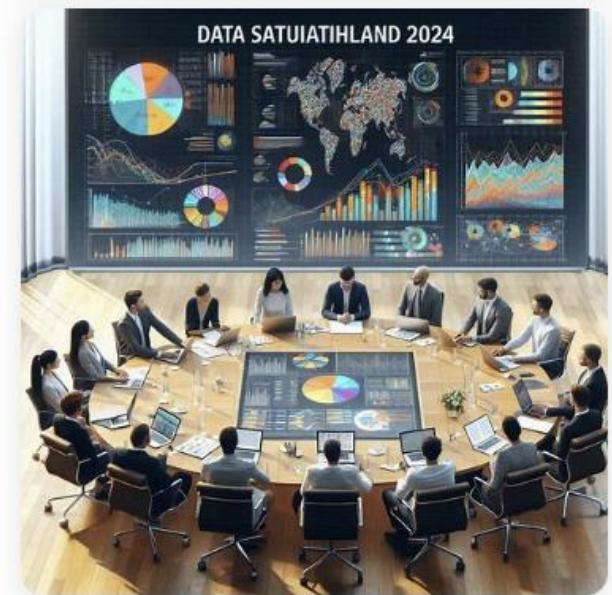


Copilot

✓ Antworten für Sie werden generiert...

Data Saturday Rheinland 2024 ist ein kostenloses eintägiges Event für Fachleute, die die Microsoft-Datenplattform und Microsoft Azure-Technologien nutzen. Diese Community-Veranstaltung bietet Inhalte zu Datenmanagement, Cloud- und Hybridarchitektur, Analytik, Business Intelligence, KI und vielem mehr ¹. Es findet bereits zum 9. Mal an der Hochschule Bonn-Rhein-Sieg statt ². Wenn du teilnehmen möchtest, kannst du dich jetzt registrieren! 😊

Weitere Informationen [1](https://datasaturdays.com) [2](https://h-brs.de) [3](https://datasaturdays.com) [+4 mehr](#)



"Data Saturday Rheinland 2024"

Who am I?

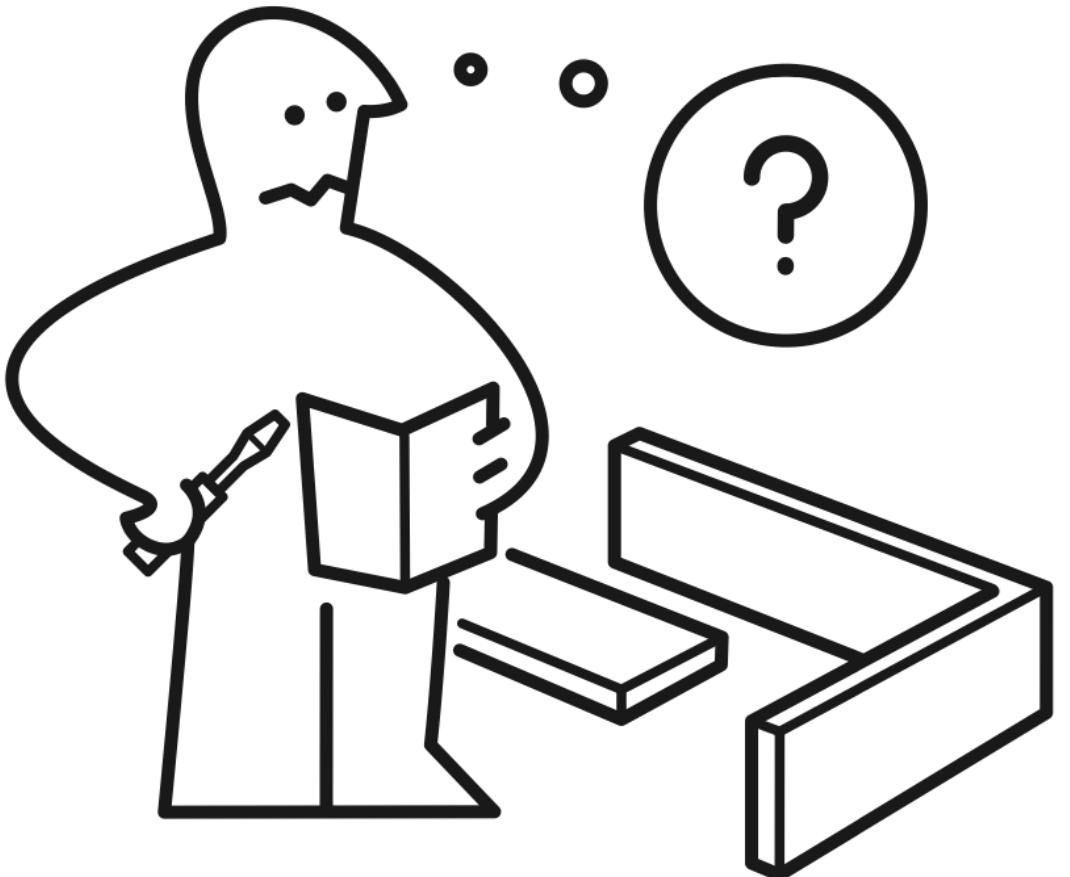
```
{  
  "name" : "Alexander Klein",  
  "description" : "Independent BI Consultant,  
  > 20 years experience of SQL Server.  
  Speaks at national and international conferences  
  & user groups about Data Warehouse, BI, Azure and AI",  
  "tags": {  
    "BI",  
    "SQL Server",  
    "Data Warehouse",  
    "Azure",  
    "AI"  
  }  
  
  "contact":  
  {  
    "email" : "a.klein@consulting-bi.de",  
    "twitter" : "@SQL_Alex",  
    "linkedin" : "https://www.linkedin.com/in/alexander-klein-ai/"  
  }  
}
```



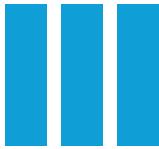
Agenda

- Generative AI
- Writing code
- Writing low/no code
- Writing text
- Generate images
- Gain new insights
- Generative AI, but ...

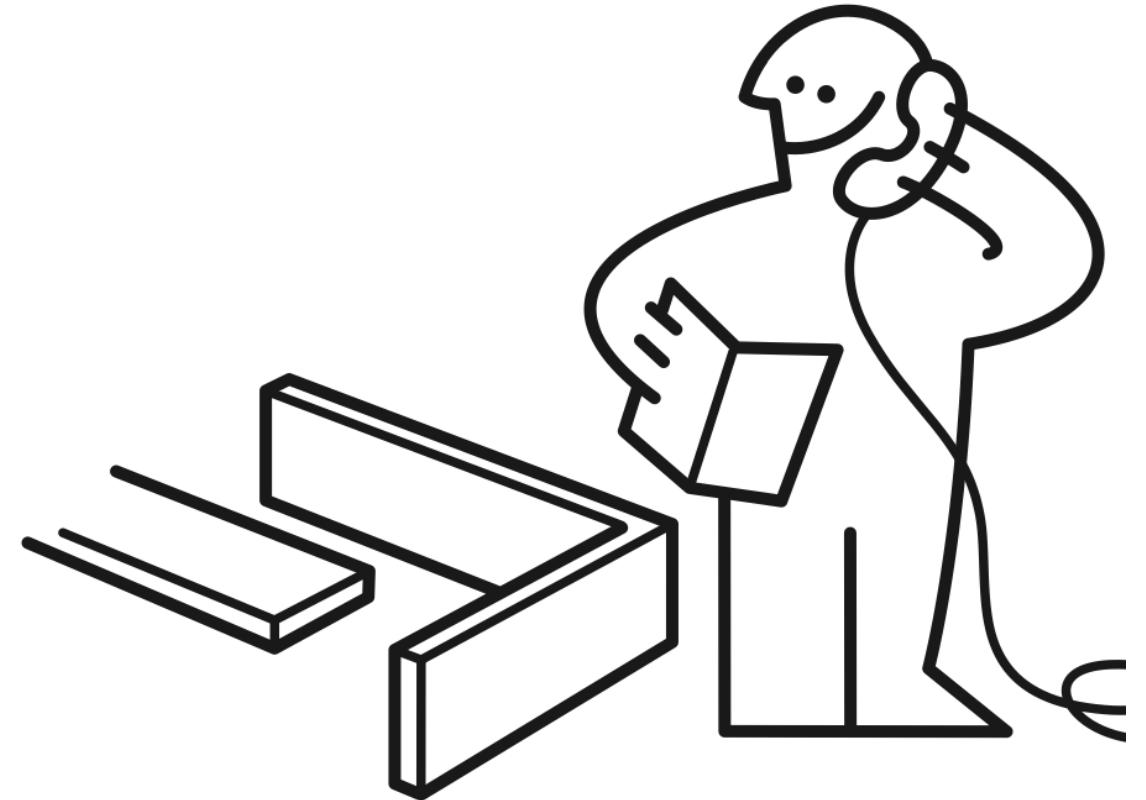




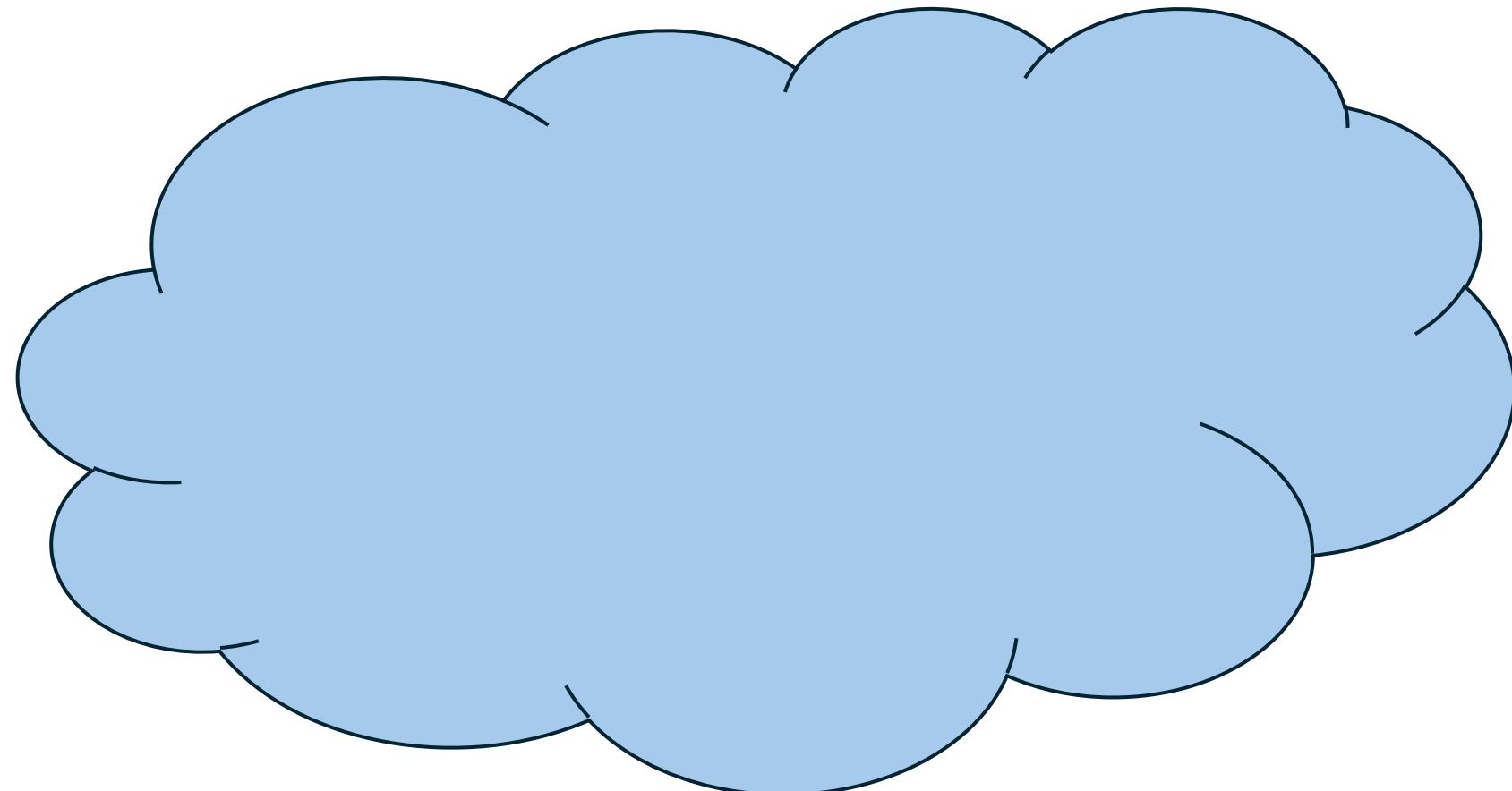
Generative AI



Generative AI



Artificial Intelligence



Machine Learning

Machine Learning

Supervised

Unsupervised

- Supervised models:
Use labeled data
- Unsupervised models:
Use unlabeled data

Deep Learning

Machine Learning

Supervised

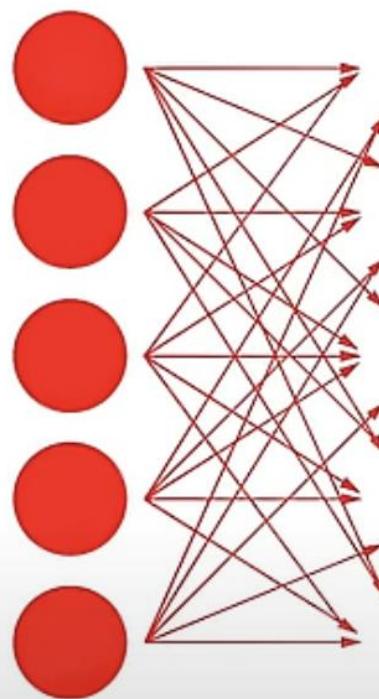
Unsupervised

Deep Learning

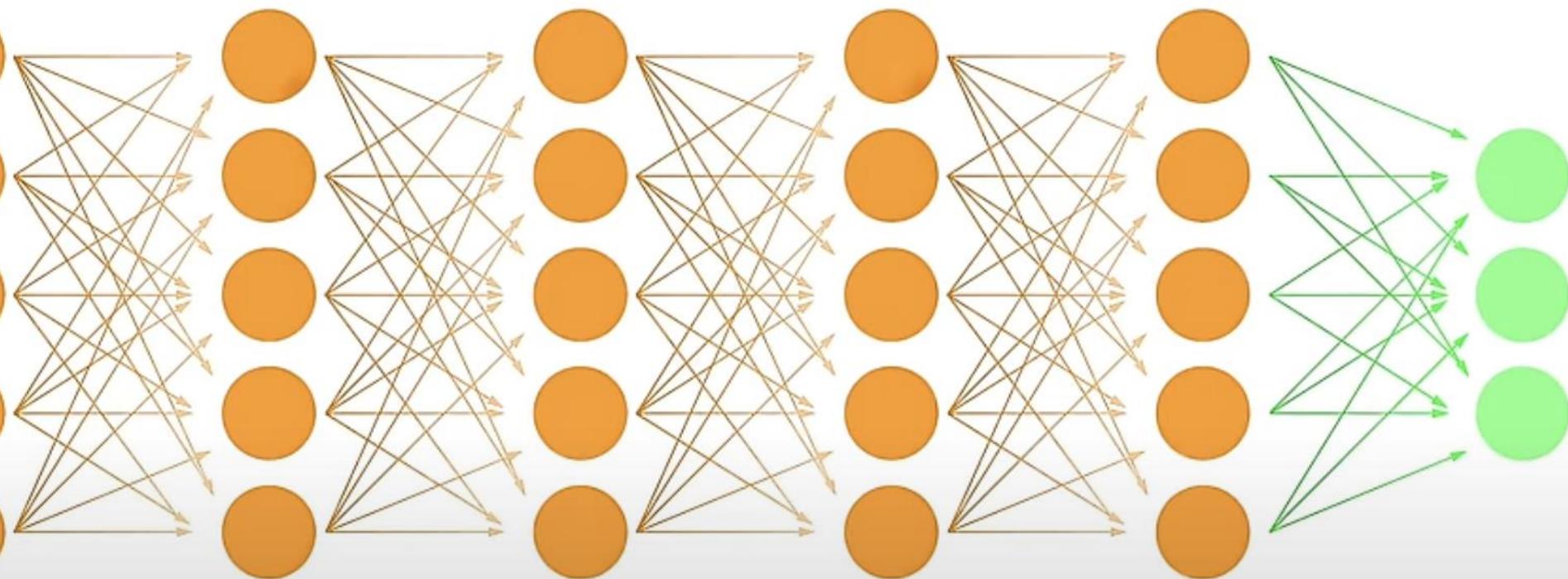
A type of machine learning that uses Artificial Neural Networks



Input layer

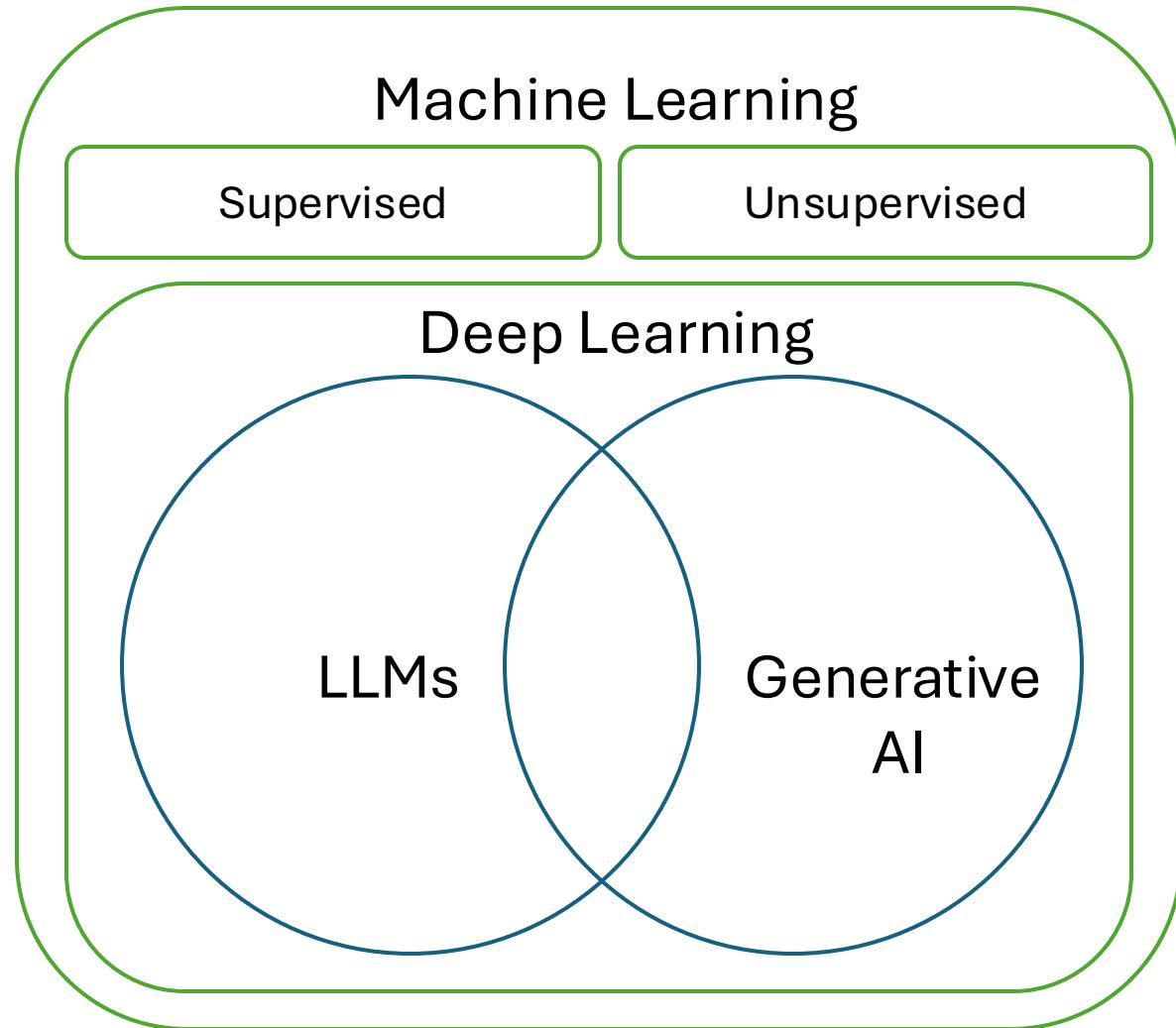


Hidden layers



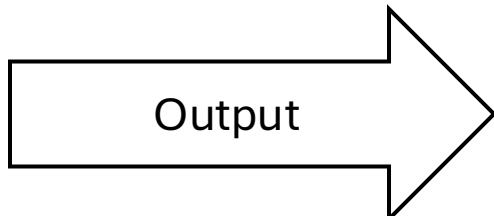
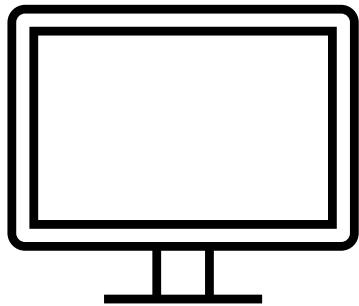
Output layer

Large Language Models (LLM)



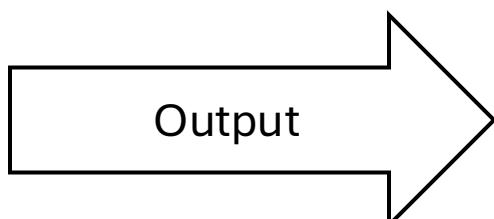
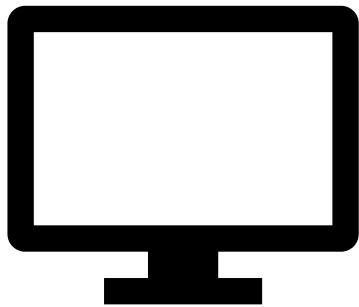
Pre-trained with a very large set of data, then fine-tuned for specific purpose

GenAI, or not GenAI: that is the question



It's NOT GenAI

- Number
- Classification
- Probability



It IS GenAI

- Natural language (text to speech)
- Image
- Audio

Generative AI Models Types

Text to text

Take natural language input and produce text output



Text to image

Trained on large set of images and generate new images



Text to video

Can generate and edit videos



Text to 3D

Can be used to produce gaming assets



Text to task

Trained to perform a specific task or action based on text input



Generative AI

Generative AI, sometimes referred to as Gen AI, is artificial intelligence (AI) that generates original content such as text, images, video, audio or software code in response to a user's prompt or request. This technology uses algorithms to generate sound, code, images, simulations, videos and text and can generate new data from previously entered information. In short, Generative AI makes it possible to automatically create new content without human intervention.



Copilot

✓ Antworten für Sie werden generiert...

Generative KI, manchmal auch als Gen AI bezeichnet, ist künstliche Intelligenz (KI), die als Antwort auf den Prompt oder die Anfrage eines Benutzers Originalinhalte wie Text, Bilder, Video, Audio oder Softwarecode generiert [1](#). Diese Technologie nutzt Algorithmen, um Ton, Code, Bilder, Simulationen, Videos und Texte zu erzeugen und kann neue Daten aus zuvor eingegebenen Informationen generieren [2](#). Kurz gesagt, Generative KI ermöglicht es, automatisch neue Inhalte zu erstellen, ohne menschliches Eingreifen [3](#).

Im Zusammenhang mit: Was ist Generative AI

Anzeigen [i](#)

Generative AI on AWS | AI Foundation Models

Deploy Scalable, Reliable, and Secure Generative AI Applications Using Amazon Web Services. AWS Offers Secure and Easy Ways To Build Generative AI Applications Quickly.

<https://aws.amazon.com/amazon/generative-ai>

Weitere Informationen [1](#) [ibm.com](#) [2](#) [bing.com](#) [3](#) [getapp.de](#) [4](#) [en.wikipedia.org](#)



Prompt Engineering/Design

Bad prompt



You

Give me an agenda for a workshop

Good prompt



You

Give me an agenda for a workshop.

task context

I'm meeting a leadership team at an aerospace consulting firm.
The goal of the workshop is figure out how they can use AI.
They are new to this. We have 8 people for 4 hours.

Good prompt



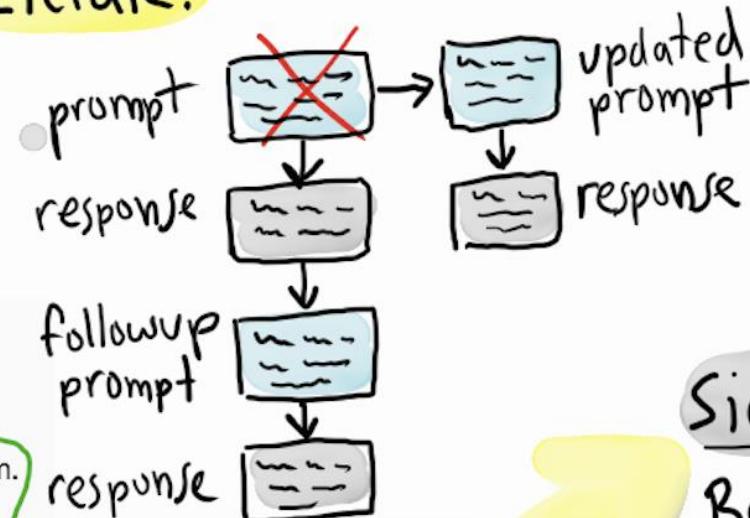
You

Give me an agenda for a workshop.

questions

Feel free to ask me any clarifying questions first.

Iterate!



Side effect

Better communication skills overall

Better prompt engineering skills = Better results

Study, Practice, Learn

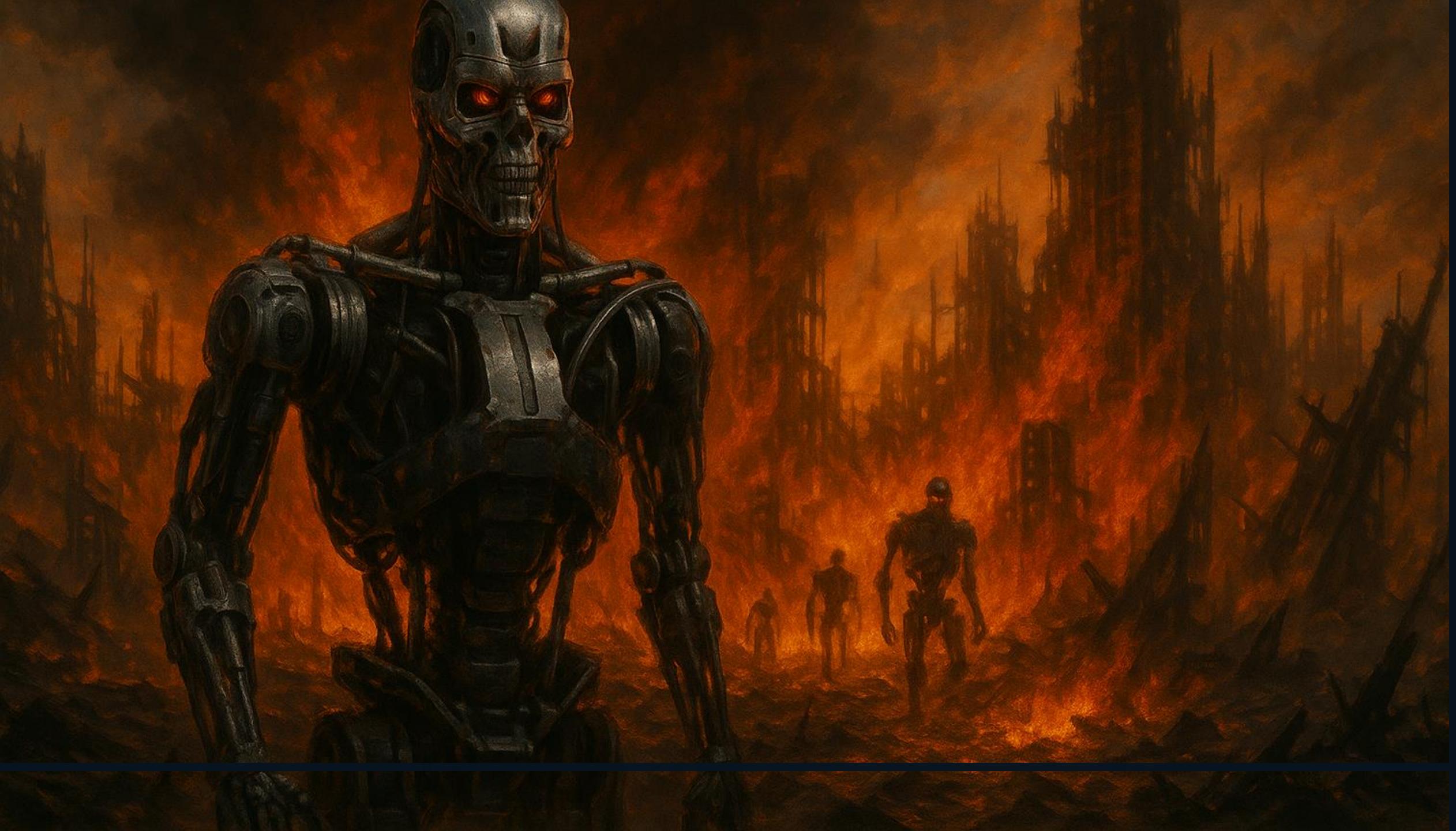
Autonomous Agents with tools



Here Be Dragons







Mindset



AI might not take your job,
but people/companies using AI will

Generative AI in a Nutshell

Prompt Engineering/Design

Bad prompt

You Give me an agenda for a workshop

Good prompt

You Give me an agenda for a workshop

task context

I'm meeting a leadership team at an aerospace consulting firm. The goal of the workshop is figure out how they can use AI. They are new to this. We have 8 people for 4 hours.

Good prompt

You Give me an agenda for a workshop. questions

questions

Free to ask me any clarifying questions first!

Autonomous Agents with tools



Here Be Dragons

Iterate!

prompt

response

followup prompt

response

updated prompt

response

side effect

Better communication skills overall

Study, Practice, Learn

questions

Free to ask me any clarifying questions first!

Using vs Building AI-powered products

The role of Humans

Is human role X still needed?

doctor, developer, lawyer, ceo, teacher, etc...

Decide what to ask and how

Evaluate results

Compensate for AI weaknesses

Provide context

Legal compliance

Data security

etc, etc...

decide what to ask and how

evaluate results

compensate for AI weaknesses

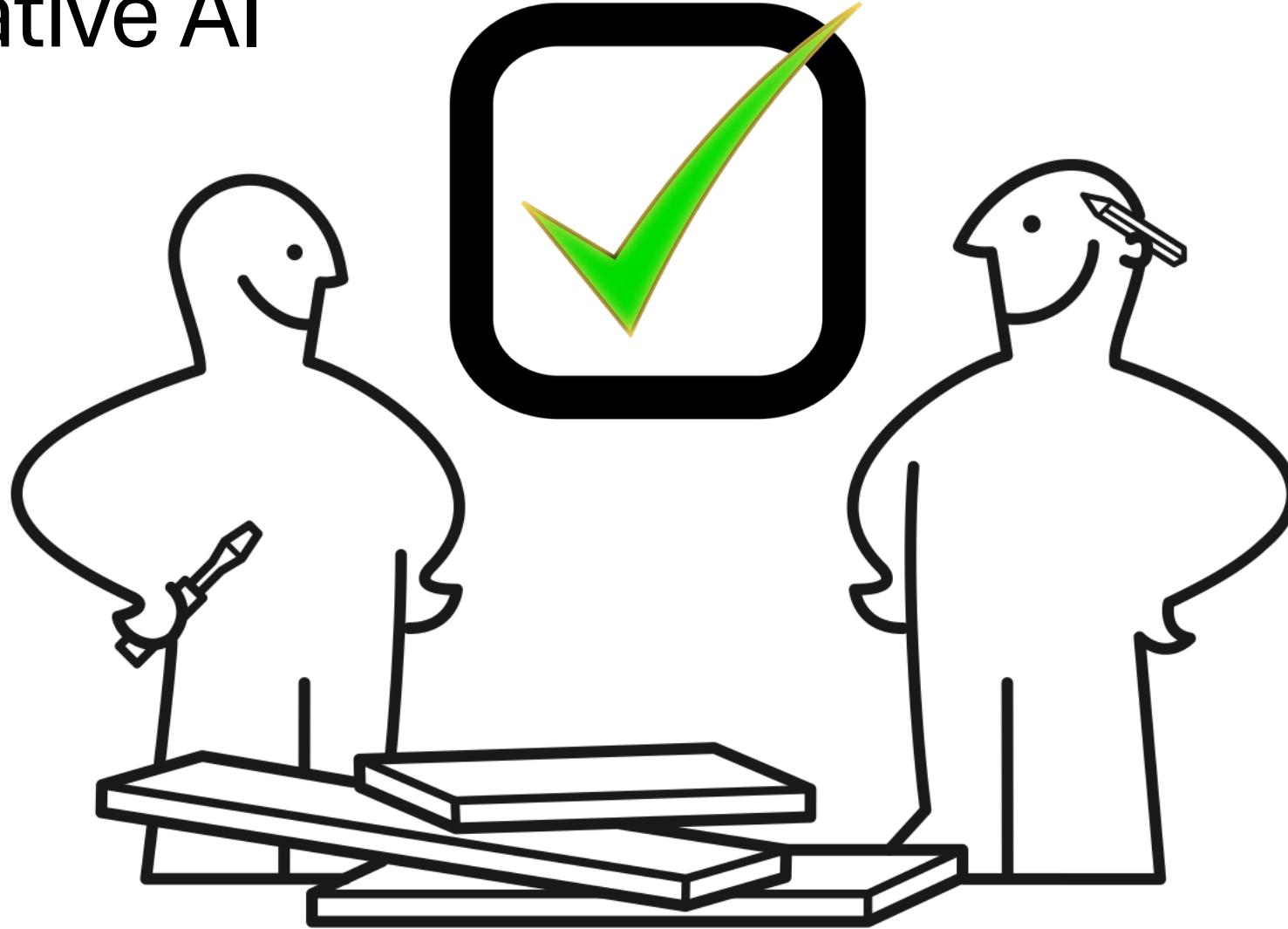
provide context

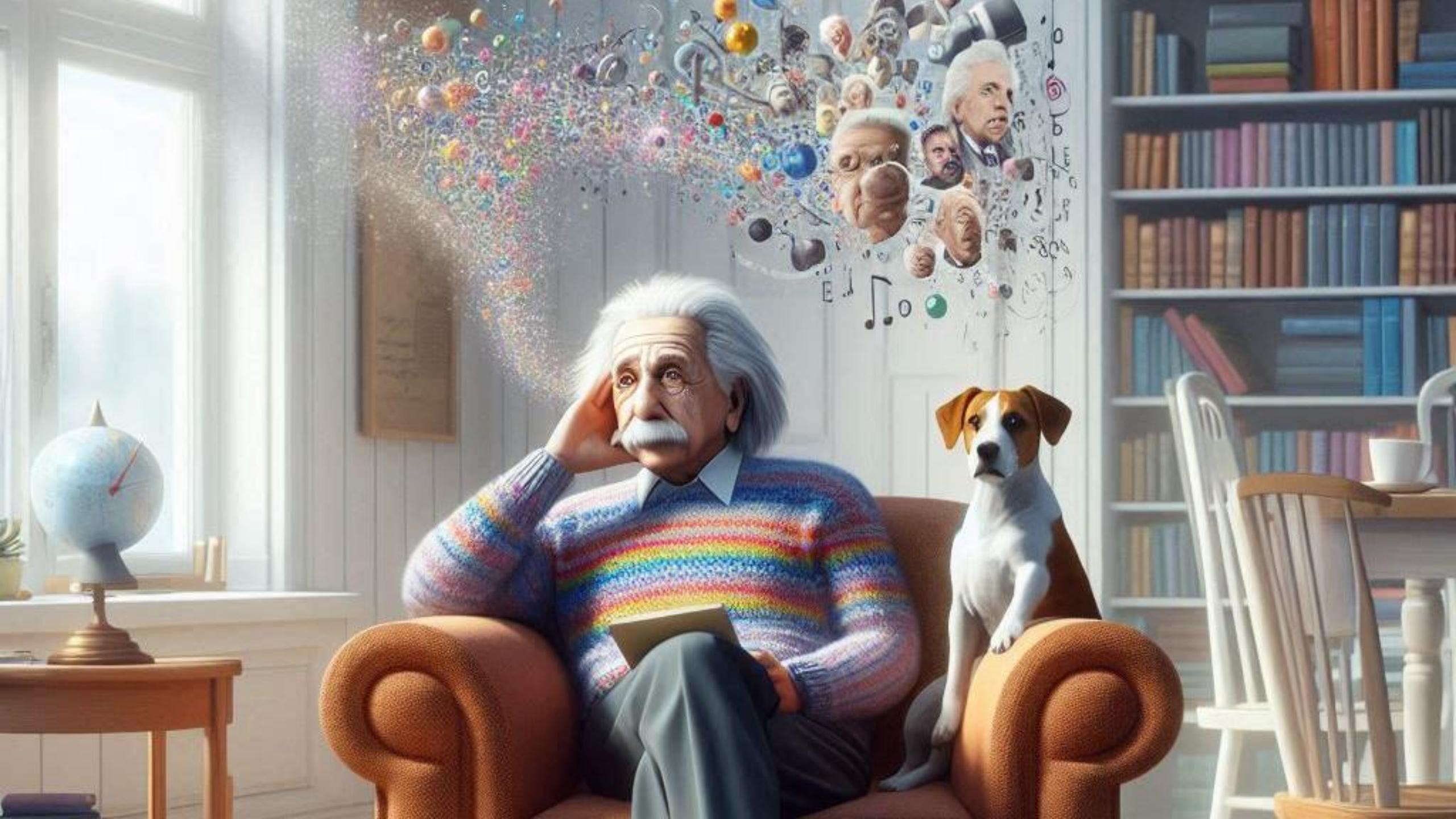
legal compliance

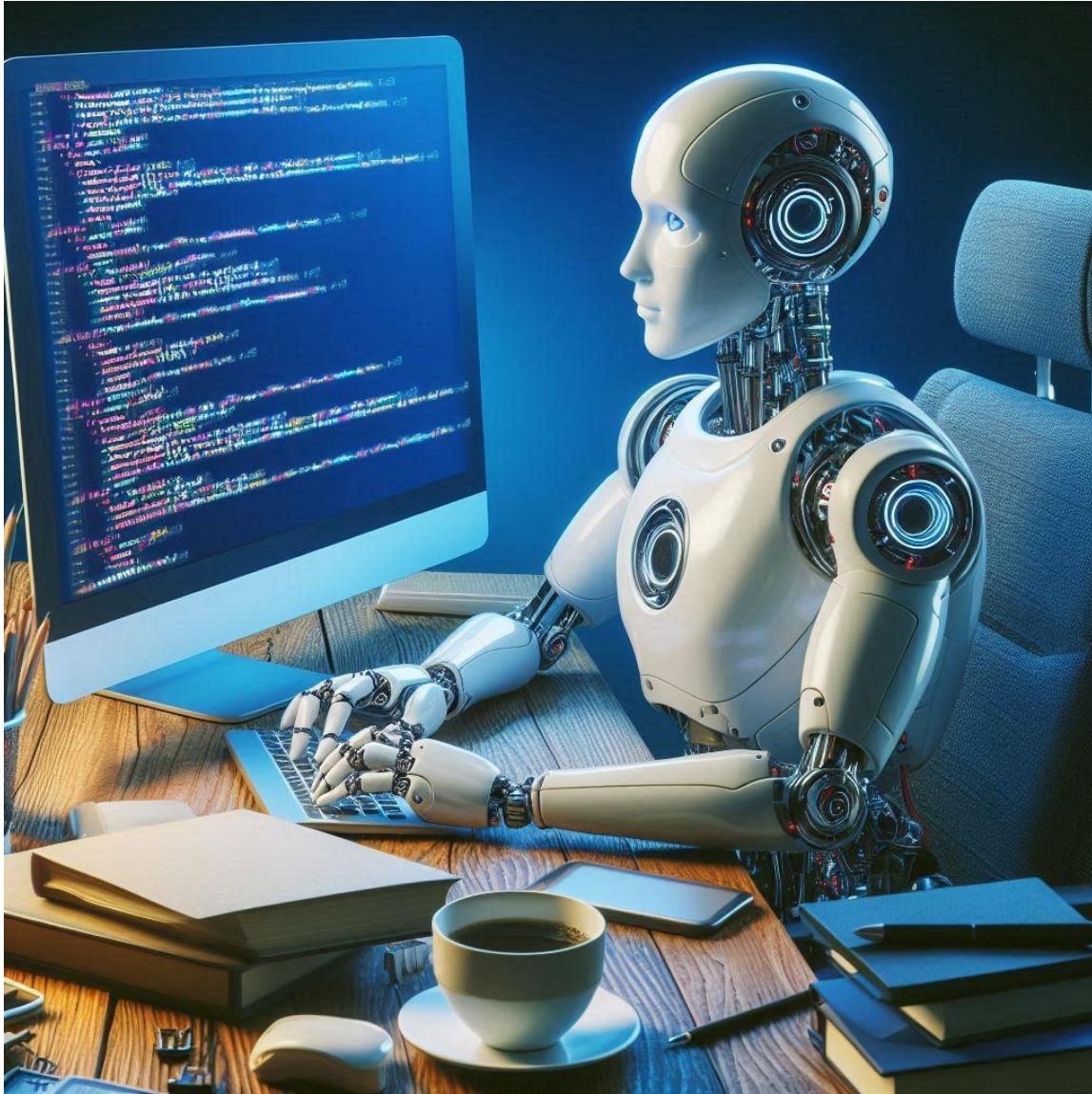
data security

etc, etc...

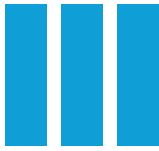
Generative AI







Writing code



Writing code



The screenshot shows a Jupyter Notebook cell with the following code:

```
fileName = [REDACTED]
filePath = [REDACTED] #ohne Dateinamen aber mit "/" am Ende
fullPath = filePath + fileName

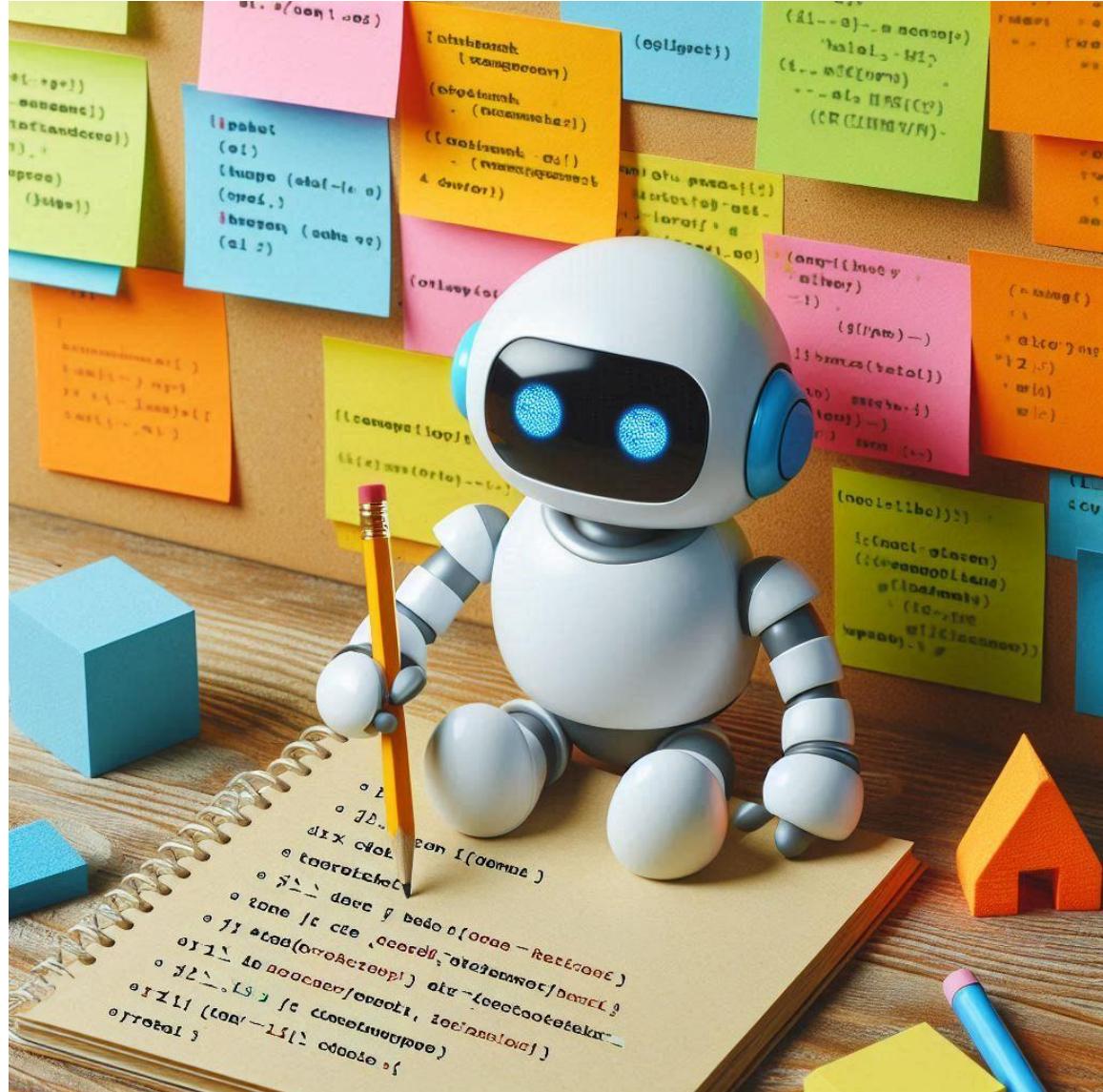
#schema = StructType([
#    StructField("_class", StringType(), True),
#    StructField("id", StructType([
#        StructField("base64", StringType(), True),
#        StructField("subType", StringType(), True),
#        StructField("comment", StringType(), True),
#        StructField("importedFromPlatform", StructType([
#            StructField("$numberLong", StringType(), True)
#        ]), True),
#        StructField("lastEditedBy", StringType(), True)
#    ]), True)
#])

#])
try:
    df = spark.read.json(fullPath, multiLine=False)
    df_ev_account = (df
                     | # DF Transformation
```

The code defines variables for file name and path, concatenates them to form a full path, and then attempts to read a JSON file at that path using Spark's `read.json` method. A comment indicates that the path should end with a slash. The code also includes a schema definition for a struct type with various fields like `_class`, `id`, and `base64`.

Writing code

```
▶ (5) Spark Jobs
  ▶ df: pyspark.sql.dataframe.DataFrame
    _class: string
    ▶ _id: struct
      $binary: string
      $type: string
      accountClassification: string
    ▶ accountIdExternal: struct
      _id: long
      checksum: long
      accountKind: string
      accountType: string
    ▶ additionalInvoiceData: array
      ▶ element: struct
        name: string
        value: string
    ▶ address: struct
      addressLine1: string
      addressLine2: string
      city: string
      countryCode: string
      firstName: string
      stateProvinceRegion: string
      surname: string
```



Writing low/no code

Writing low/no code

The screenshot displays the Microsoft Copilot AI interface, which is designed to help users build applications and automate tasks using natural language descriptions.

Describe your copilot to create it

Helpdesk, Expense tracking, HR and benefits

Use everyday words to describe what your copilot should do

This AI-powered feature above is currently in preview. [See terms](#) and [supplemental terms](#)

Let's build an app. What should it do?

Collect RSVPs, Track sales leads, List inventory, Manage inspections

Use everyday words to describe what your app should collect, track, list, or manage ...

This feature uses generative AI. [See terms](#)

Let's automate something. What should it do?

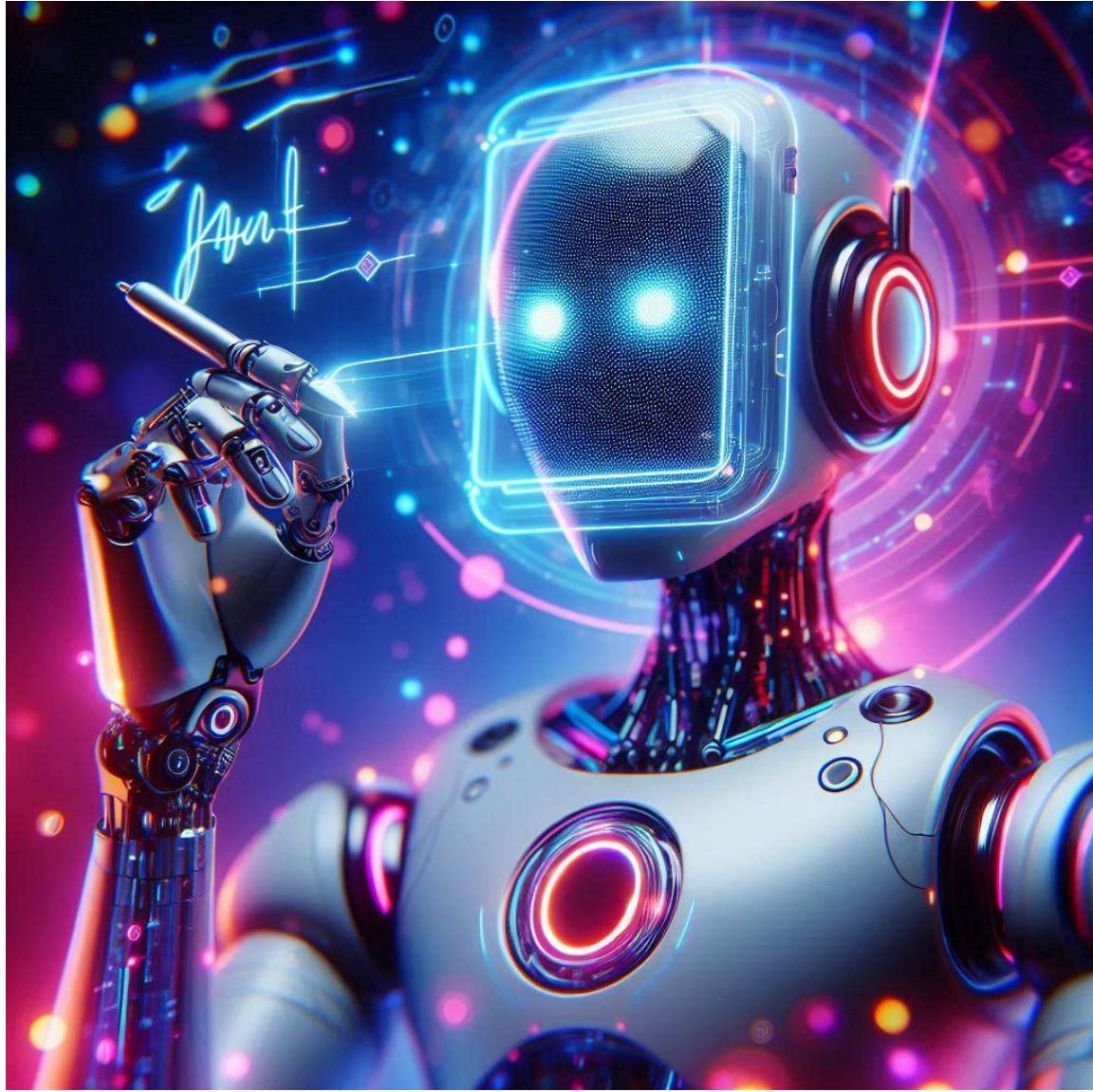
Get started by selecting an example or describing your own automation idea.

Every month, copy all files from OneDrive folder to another OneDrive folder
Copy all rows from an Excel file to another excel file with a click of a button
When a new item is created in SharePoint, send me an email

Describe in detail how you want your automation to work

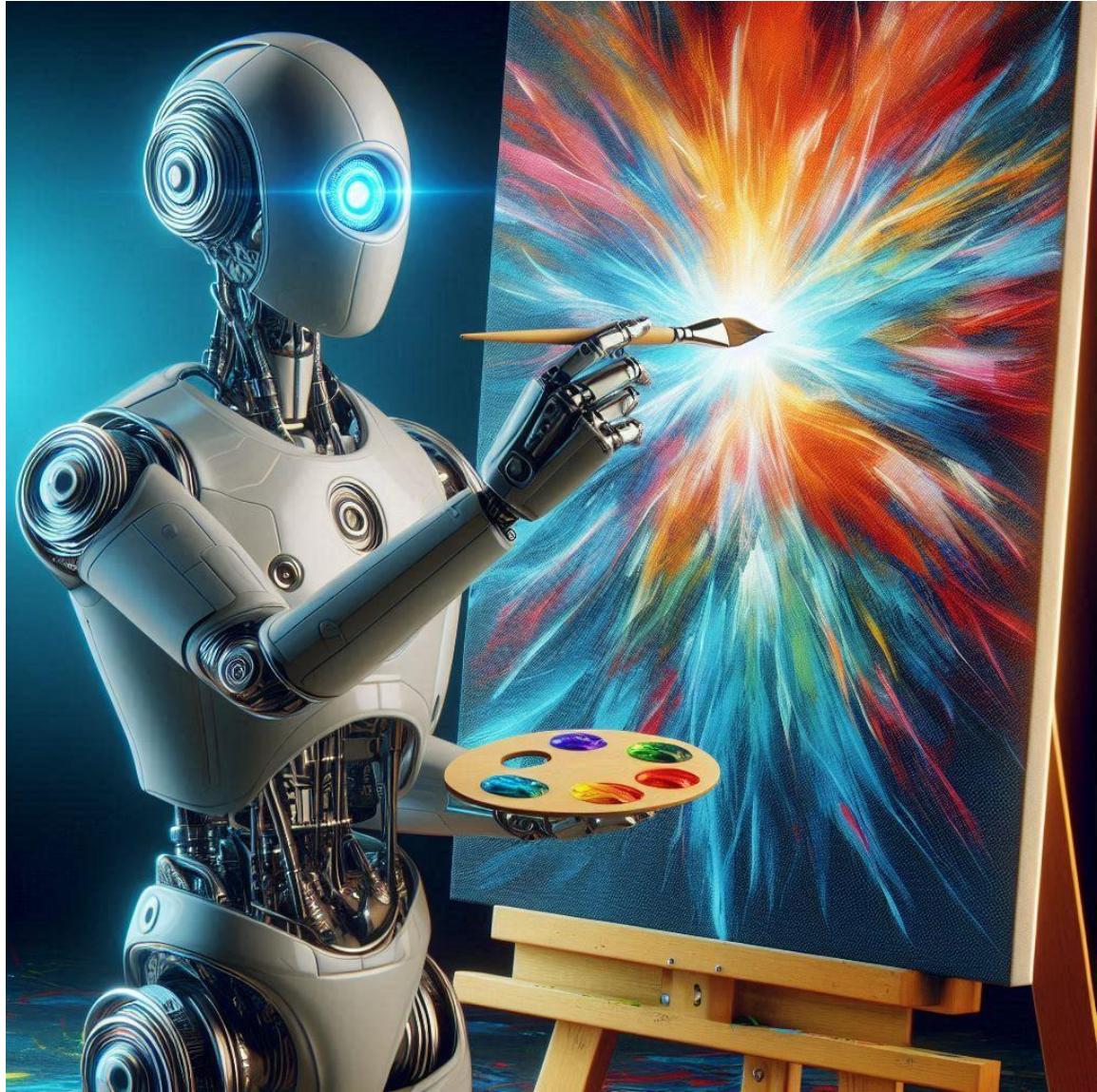
Generate

Copilot is new technology that is still being developed. It is optimized for use with English language and has limited support with other languages. As such, parts of it may appear in English rather than your preferred language. [Learn more](#)

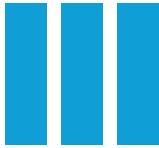


Writing text





Generate
images



Data Saturday #50



Sie

male mir ein Bild zum 50. Geburtstag des Data Saturday

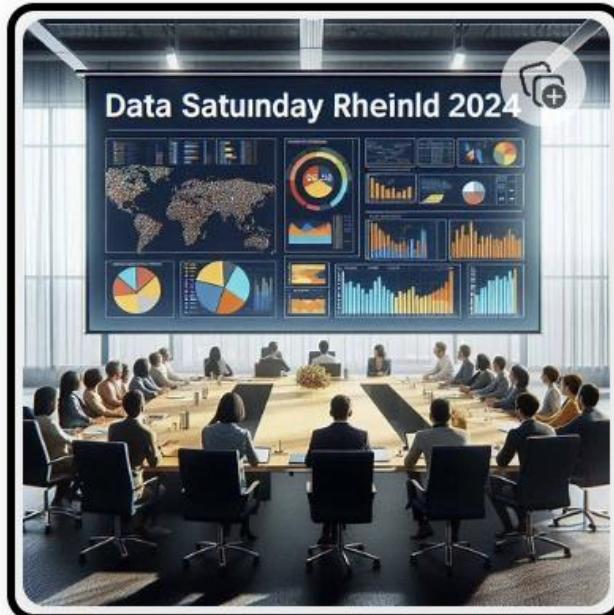


Copilot

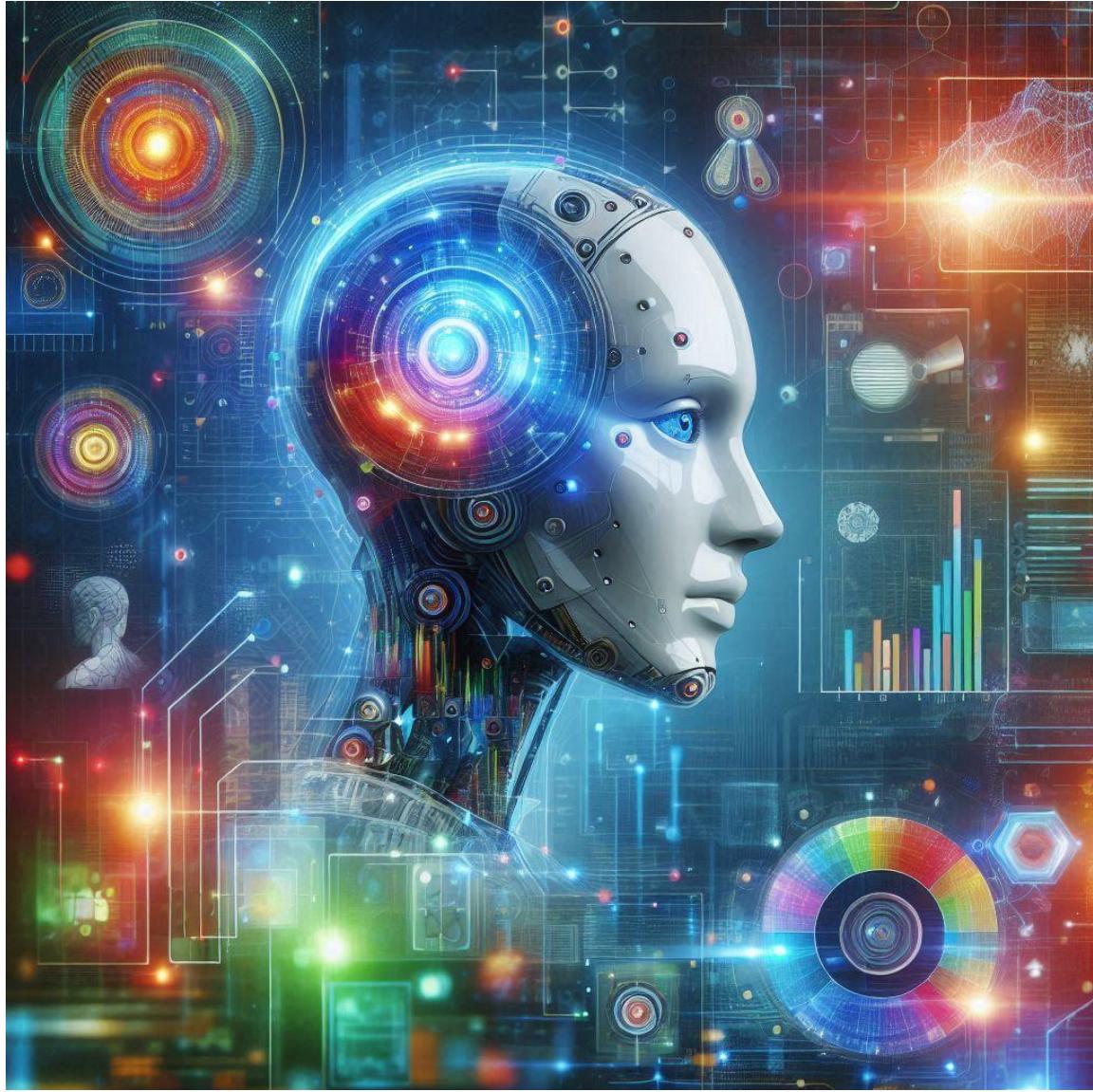
✓ Antworten für Sie werden generiert...

Data Saturday Rheinland 2024 ist ein kostenloses eintägiges Event für Fachleute, die die Microsoft-Datenplattform und Microsoft Azure-Technologien nutzen. Diese Community-Veranstaltung bietet Inhalte zu Datenmanagement, Cloud- und Hybridarchitektur, Analytik, Business Intelligence, KI und vielem mehr¹. Es findet bereits zum 9. Mal an der Hochschule Bonn-Rhein-Sieg statt². Wenn du teilnehmen möchtest, kannst du dich jetzt registrieren! 😊

Weitere Informationen [1 datasaturdays.com](https://datasaturdays.com) [2 h-brs.de](https://h-brs.de) [3 datasaturdays.com](https://datasaturdays.com) +4 mehr



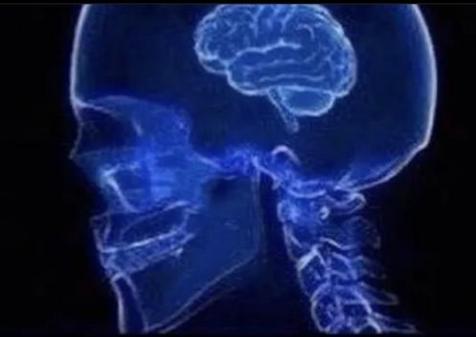
"Data Saturday Rheinland 2024"



Gain new
insights



Sorry I forgot



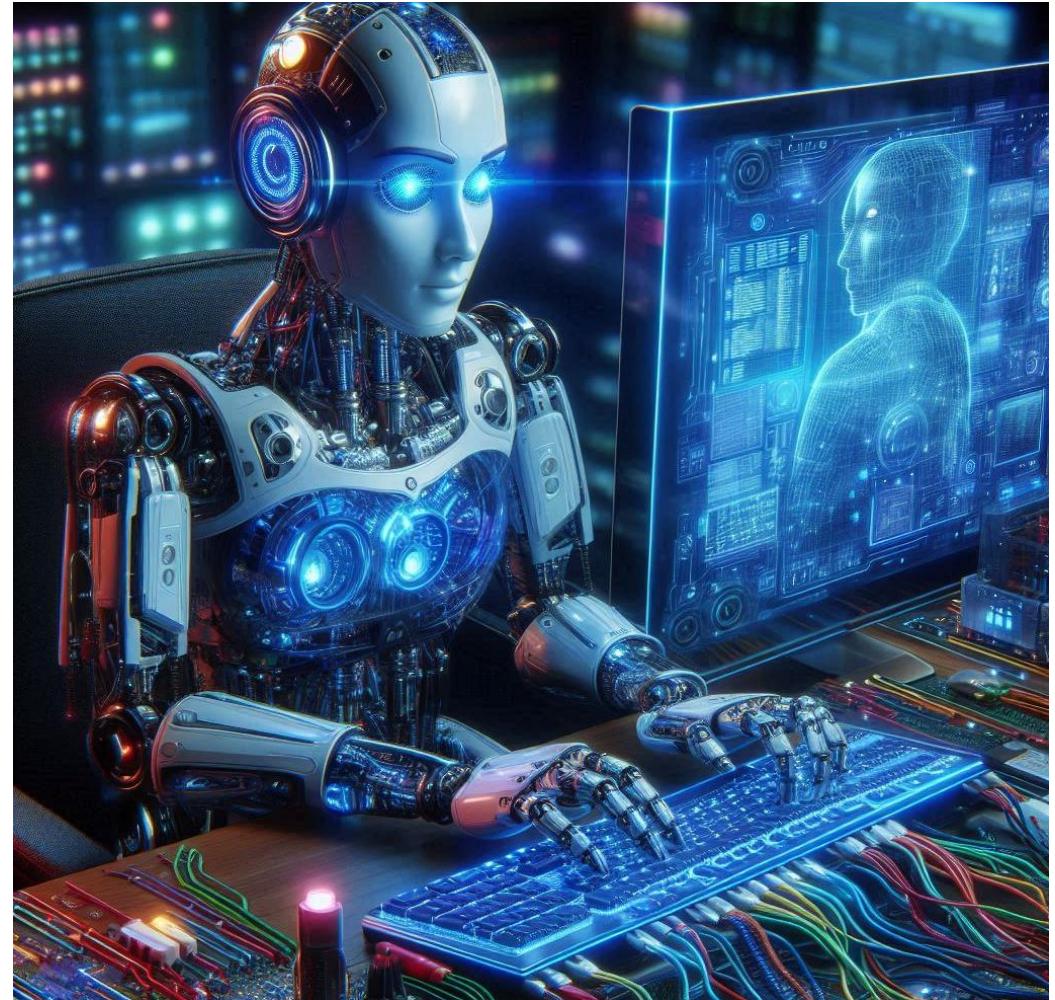
I 4got



Generative AI ...

Generative AI yes, but ...

- Data protection (company data)
- Integration into existing software (e.g. Office, Teams, Browser, ...)
- Human as limitation (Albert Einstein)
- Test each new version
- Don't blindly trust the AI !
→ Always critically evaluate the proposal



**ONLY WAY TO KEEP
AI FROM TAKING YOUR JOB**

**IS TO USE AI TO
DO YOUR JOB BETTER**