

Session B4

AI trifft auf BI: Wie ChatGPT und andere KI-Tools die BI-Entwicklung vereinfachen



Elgar Brunott

Principal Consultant Data & Analytics
ebrunott@it-logix.ch



Severin Leuenberger



Principal Consultant Data & Analytics
sleuenberger@it-logix.ch



Pascal Schär

Senior Consultant Data & Analytics
pschaer@it-logix.ch

Stellen Sie Fragen in der virtuellen Plattform

- 1 Loggen Sie sich auf der virtuellen Plattform mit **Ihrem persönlichen Link** ein (aus Bestätigungs-Mail MAKE BI)
- 2 Kicken Sie in der Lobby auf der **Session B4** auf den Button „**Teilnehmen**“

- 3 Kicken Sie in der **Session B4** auf der rechten Seite in der Navigation auf „**Q&A**“
MOBILE: Klicken Sie der Navigation unten zuerst auf „**Interaktion**“, danach auf das **Sprechblasen-Symbol**

- 4 Geben Sie im entsprechenden Feld Ihre Frage ein und schicken Sie diese ab, indem Sie auf das **Pfeil-Symbol** klicken

Die Referenten werden am Ende Ihrer Präsentation vereinzelt auf Fragen eingehen. Nicht beantwortete Fragen können gerne bei den diversen Networking-Möglichkeiten mit den entsprechenden Referenten aufgegriffen werden.

● chatgpt
Suchbegriff



● switzerland
Suchbegriff



● COVID-19
Suchbegriff



Weltweit ▼

01.01.20 bis 24.08.23 ▼

Alle Kategorien ▼

Websuche ▼

Interesse im zeitlichen Verlauf



● chatgpt

Suchbegriff

⋮

● switzerland

Suchbegriff

⋮

● COVID-19

Suchbegriff

⋮

Weltweit ▾

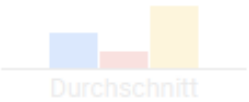
01.01.20 bis 24.08.23 ▾

Alle Kategorien ▾

Websuche ▾

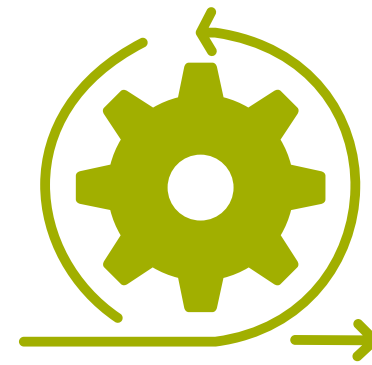


Interesse im zeitlich





Wie uns ChatGPT im BI-Umfeld unterstützen kann – 5 Anwendungsbeispiele



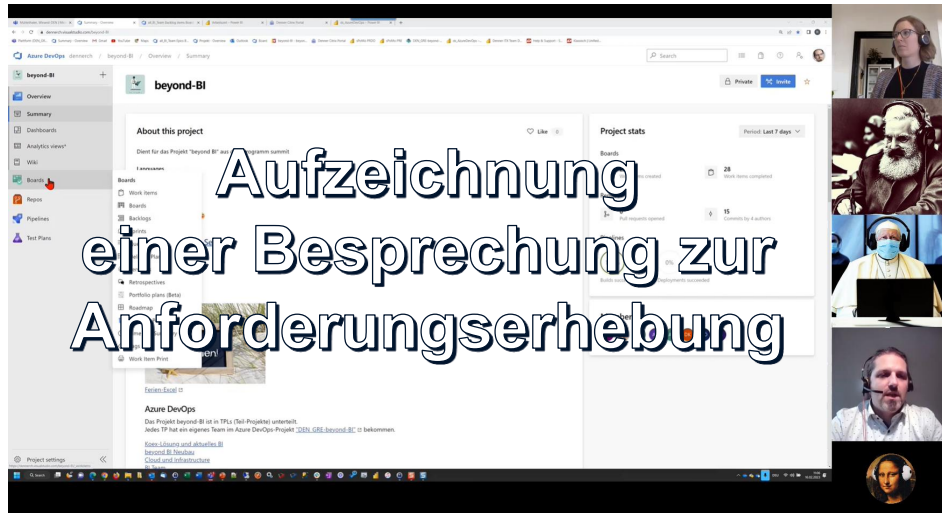


Beispiel 1: Anforderungsanalyse – Ausgangslage und Herausforderung

The screenshot shows the Azure DevOps interface for a project named 'beyond-BI'. The left sidebar contains navigation links: Overview, Summary, Dashboards, Analytics views, Wiki, Boards, Repos, Pipelines, and Test Plans. The 'Boards' link is highlighted, and a dropdown menu is open, showing options like Work items, Boards, Backlogs, Sprints, Queries, Delivery Plans, Personas, Retrospectives, Portfolio plans (Beta), Roadmap, Planning Poker, Time Log Summary, Tags, and Work Item Print. The main content area is titled 'beyond-BI' and includes a section 'About this project' with the text 'Dient für das Projekt "beyond BI" aus dem Programm summit'. Below this is a 'Project stats' section for the 'Last 7 days' period, showing 22 work items created, 28 work items completed, 0 pull requests opened, and 15 commits by 4 authors. The 'Pipelines' section shows 100% builds succeeded and 0% deployments succeeded. The 'Members' section lists 51 team members. On the right side of the screen, there is a vertical stack of four video call windows. The top window shows a woman with glasses and a headset. The second window shows a man with a beard and a headset. The third window shows a man with a blue face mask and a headset. The bottom window shows a man with a headset. At the very bottom of the video call stack is a small circular icon of the Mona Lisa.



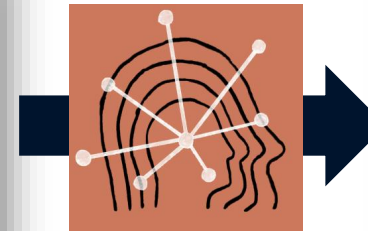
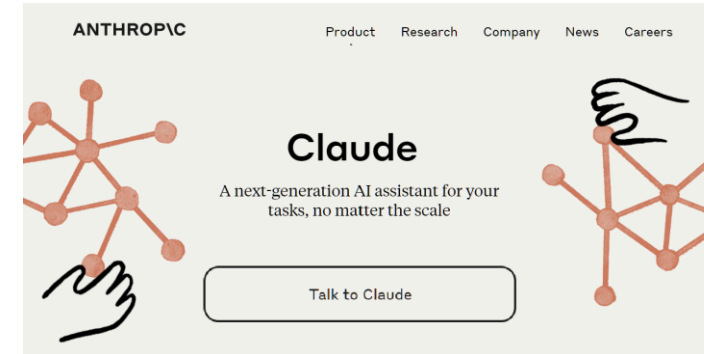
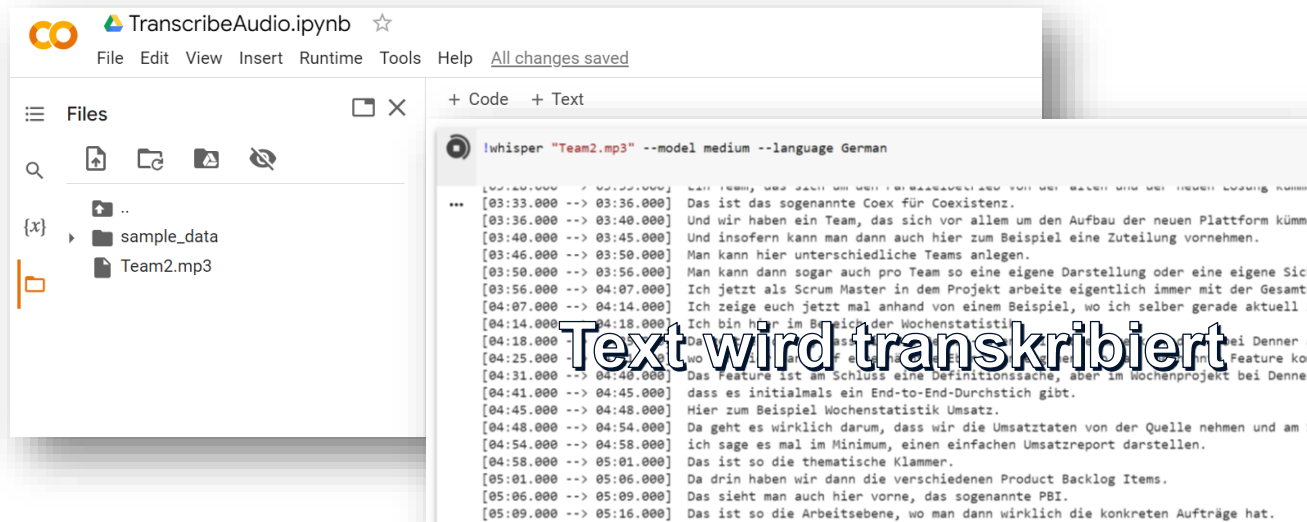
Beispiel 1: Anforderungsanalyse – Ausgangslage und Herausforderung



1. Video sichten
2. Anforderungen notieren
3. Anforderungen zusammenfassen



Beispiel 1: Anforderungsanalyse – Vorgehen



- Raphael von IT-Logix gibt eine Demo und Einführung in Azure DevOps, wo IT-Logix seit 1,5 Jahren ein BI-System aufbaut.

- Azure DevOps hat verschiedene Komponenten. Der Fokus liegt auf Anforderungen als Backlog Items verwaltet und die Teams können

- Raphael zeigt die Backlog-Ansicht mit einer hierarchischen Struktur. Diese dienen als eigentliche Beschreibung und Acceptance Criteria. Updates zu den Items ausgetauscht werden.

- Über ein Board hat man den Überblick über den Status der Tickets.

Beispiel 2A: Testdaten generieren – Ausgangslage

- › oft ist die Struktur der Quelldaten bekannt, aber die Daten selbst (noch) nicht verfügbar («1000 Gründe»)
- › um weiter arbeiten zu können, macht es aber häufig Sinn, konkret mit Daten arbeiten zu können (ist sonst wie Trockenübungen beim Schwimmen)
- › ChatGPT kann Daten generieren anhand einer vorgegebenen Struktur

Beispiel 2A: Testdaten generieren – Vorgehen

The screenshot displays the Microsoft Fabric workspace interface. The top navigation bar is green and contains the text "Testdaten generieren" and "Saved". A search bar is located on the right. The left sidebar shows various icons for navigation, including Home, Create, Browse, OneLake data hub, Monitoring hub, Workspaces, and a specific workspace named "MAKE BI 2023...". The main area has a title bar with "Home", "Edit", "Run", "Data", and "View" tabs. Below this is a toolbar with icons for file operations, a "Run all" button, a "Stop session" button, a "Language" dropdown set to "PySpark (Python)", and an "Open in VS Code" button. The main content area is titled "Beispiel2: Testdaten generieren". It shows two execution blocks. The first block, labeled [5], contains the command `!pip install openai` and a status message: "✓ -Command executed in 2 sec 706 ms by Pascal Schär on 11:49:51 AM, 7/03/23". Below this, it says "Output is hidden". The second block, labeled [3], contains a Python script for generating completions using the OpenAI API. The script defines a `get_completion` function that takes a prompt and a model name, and returns the generated content. The status message for this block is: "✓ -Command executed in 398 ms by Pascal Schär on 11:47:14 AM, 7/03/23".

Testdaten generieren | Saved

Search

Home Edit Run Data View

Run all Stop session Language PySpark (Python) Open in VS Code

Beispiel2: Testdaten generieren

```
1 !pip install openai
```

[5] ✓ -Command executed in 2 sec 706 ms by Pascal Schär on 11:49:51 AM, 7/03/23

Output is hidden

```
1 import openai
2 import pandas as pd
3
4 openai.api_key = 'sk-1jprg2e1H8ag1F6CZ9nQ708J6w70x1M677heW7u444FpJ1Tn'
5
6 def get_completion(prompt, model="gpt-3.5-turbo"):
7     messages = [{"role": "user", "content": prompt}]
8     response = openai.ChatCompletion.create(
9         model=model,
10        messages=messages,
11        temperature=0, # this is the degree of randomness of the model's output
12    )
13    return response.choices[0].message["content"]
```

[3] ✓ -Command executed in 398 ms by Pascal Schär on 11:47:14 AM, 7/03/23

Beispiel 2A: Testdaten generieren – Prompt

```
1 pre_prompt = f"""
2 Create an example dataset with 100 records of patient data from a Swiss University Hospital containing the following fields:
3 """
4
5 data_template = f"""
6 Patient_ID: integer starting at 1 counting up,
7 Citizenship: two letter ISO country code CH or FR or DE or IT,
8 First_Name: random first name matching citizenship,
9 Last_Name: random last name matching citizenship,
10 Email: private or professional email address from random domain or email provider,
11 Birthdate,
12 Age,
13 Gender_Code,
14 AHV_Number: random Swiss social security number in the format 756.xxxx.xxxx.xx for Swiss citizen (Citizenship CH) or "n/a" otherwise,
15 Main_Diagnosis: random ICD-10 code matching the persons age and gender.
16 """
17
18 post_prompt = f"""
19 Format it as .csv and give me just the header and data and suppress any other text.
20 """
21
22
23 response = get_completion(pre_prompt + data_template + post_prompt)
24 print(response)
```

✓ -Command executed in 2 min 11 sec 368 ms by Pascal Schär on 12:11:24 PM, 7/03/23

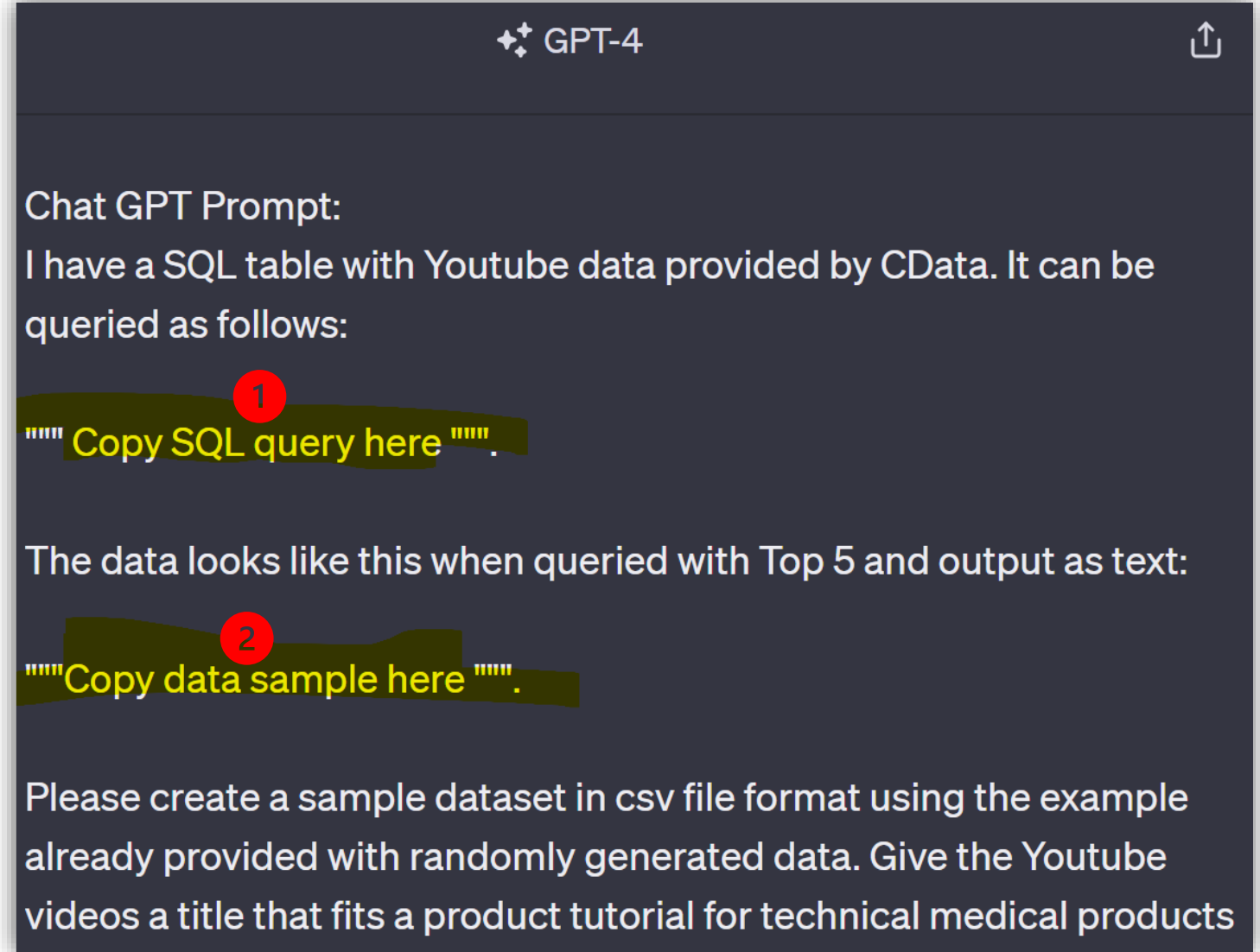
PySpark (Python) ▾

Beispiel 2A: Testdaten generieren – Ergebnis

```
Patient_ID,Citizenship,First_Name,Last_Name,Email,Birthdate,Age,Gender_Code,AHV_Number,Main_Diagnosis
1,CH,Anna,Müller,anna.muller@gmail.com,1985-03-12,36,F,756.1234.5678.90,A09,Gastroenteritis and colitis
2,FR,Étienne,Dupont,etienne.dupont@hotmail.com,1972-09-25,49,M,n/a,I10,Essential (primary) hypertension
3,DE,Michael,Schmidt,michael.schmidt@yahoo.com,1990-07-05,31,M,n/a,S72.001A,Femoral neck fracture
4,IT,Giulia,Rossi,giulia.rossi@gmail.com,1988-11-18,33,F,n/a,M54.5,Low back pain
5,CH,Thomas,Keller,thomas.keller@bluewin.ch,1976-06-30,45,M,756.9876.5432.10,J45.909,Asthma, unspecified
6,FR,Marie,Lefèvre,marie.lefevre@hotmail.com,1995-02-14,26,F,n/a,R10.4,Other and unspecified abdominal pain
7,DE,Andreas,Wagner,andreas.wagner@yahoo.com,1982-04-03,39,M,n/a,I25.10,Atherosclerotic heart disease
8,IT,Luca,Romano,luca.romano@gmail.com,1978-08-22,43,M,n/a,K21.9,Gastro-esophageal reflux disease
9,CH,Sandra,Baumann,sandra.baumann@bluewin.ch,1993-01-07,28,F,756.3456.7890.12,M54.16,Radiculopathy
10,FR,Julie,Moreau,julie.moreau@hotmail.com,1980-12-01,41,F,n/a,I50.9,Heart failure, unspecified
11,DE,Stefan,Müller,stefan.muller@yahoo.com,1974-05-19,47,M,n/a,S42.001A,Fracture of unspecified bone
12,IT,Marco,Rossi,marco.rossi@gmail.com,1991-10-03,30,M,n/a,M54.41,Lumbago with sciatica, right side
13,CH,Monika,Schneider,monika.schneider@bluewin.ch,1987-07-16,34,F,756.2345.6789.01,J44.9,Chronic obstructive pulmonary disease
14,FR,Philippe,Dubois,philippe.dubois@hotmail.com,1979-03-29,42,M,n/a,I48.91,Unspecified atrial fibrillation
15,DE,Christine,Fischer,christine.fischer@yahoo.com,1996-09-12,25,F,n/a,S72.002A,Fracture of unspecified bone
16,IT,Francesca,Romano,francesca.romano@gmail.com,1983-11-25,38,F,n/a,K29.70,Alcoholic gastritis
17,CH,Markus,Keller,markus.keller@bluewin.ch,1977-06-09,44,M,756.8765.4321.09,J45.40,Moderate asthma
18,FR,Isabelle,Lefèvre,isabelle.lefevre@hotmail.com,1992-01-23,29,F,n/a,R10.9,Unspecified abdominal pain
19,DE,Andrea,Wagner,andrea.wagner@yahoo.com,1989-03-07,32,F,n/a,I25.110,Atherosclerotic heart disease
20,IT,Roberto,Rossi,roberto.rossi@gmail.com,1975-08-01,46,M,n/a,K21.0,Gastro-esophageal reflux disease
```

Beispiel 2B: Testdaten generieren

- › In gewissen Situationen können Daten zur Verfügung stehen, jedoch nicht in der für die Entwicklung benötigten Menge oder Diversität.
- › um weiter arbeiten zu können, macht es Sinn, konkret mit Daten arbeiten zu können
- › ChatGPT kann anhand von Beispielen synthetische Datensätze mit den gewünschten Eigenschaften generieren



Youtube Data / benötigte Testdaten erzeugen

SQLQuery1.sql - SA...n.leuenerberger (63))

SELECT top 5 [Id]

[As_Of_Date]

[Title]

[PublishedAt]

[Duration]

[ViewCount]

[LikeCount]

[DislikeCount]

[CommentCount]

[EmbedHtml]

[psa_load_id]

[psa_record_source]

[psa_load_date]

[hash_value]

[dss_update_time]

[dss_create_time]

FROM [DWH_DEV].[yot_out].[psa_YOT_Videos_cur]

order by As_Of_Date desc

107 %

Results Messages

	Id	As_Of_Date	Title	PublishedAt	Duration	ViewCount	LikeCount	DislikeCount	CommentCount	EmbedHtml	psa_load_id
1	_2JXxB1378	2023-08-23		2017-11-13 15:47:15.0000000 +01:00	PT29S	226	2	0	0	<iframe width="480" height="270" src="//www.yout...	0
2	_7e5gamO3nA	2023-08-23		2011-06-06 16:17:24.0000000 +02:00	PT3M15S	352	1	0	0	<iframe width="480" height="270" src="//www.yout...	0
3	_adoo5PLtE	2023-08-23		2018-11-08 10:26:34.0000000 +01:00	PT37S	86	0	0	0	<iframe width="480" height="270" src="//www.yout...	0
4	_as1hv2yjc	2023-08-23		2020-06-14 00:26:36.0000000 +02:00	PT51M10S	1926	24	4	0	<iframe width="480" height="270" src="//www.yout...	0
5	_BDiejswf9Y	2023-08-23		2022-11-25 20:36:37.0000000 +01:00	PT15S	1	0	0	0	<iframe width="480" height="270" src="//www.yout...	0



Chat GPT Prompt:

I have a SQL table with Youtube data provided by CData. It can be queried as follows:

```

=====
SELECT [Id]
      ,[As_Of_Date]
      ,[Title]
      ,[PublishedAt]
      ,[Duration]
      ,[ViewCount]
      ,[LikeCount]
      ,[DislikeCount]
      ,[CommentCount]
      ,[EmbedHtml]
      ,[psa_load_id]
      ,[psa_record_source]
      ,[psa_load_date]
      ,[hash_value]
      ,[dss_update_time]
      ,[dss_create_time]
FROM [DWH_DEV].[yot_out].[psa_YOT_Videos_cur]
=====

```

The data looks like this when queried with Top 5 and output as text:

```

=====
_2IJXB1378;2023-08-23;Vielen Dank für Dein Engagement!;2017-11-13 15:47:15.0000000
+01:00;PT29S;226;2;0;0;"<iframe width=""480"" height=""270""
src=""//www.youtube.com/embed/_2IJXB1378"" frameborder=""0"" allow=""accelerometer; a
clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"" allowfullscreer
</iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08-24
03:00:13.867;80AED19460A9B2C83AA13CDBA524E0A05146C4547F494B76FB530CBADA
08-24 03:00:13.990;2023-08-24 03:00:13.990
_7e5gamO3nA;2023-08-23;Interview avec Céline Riard, Secrétaire divisionnaire de jeunesse
(SR).mp4;2011-06-06 16:17:24.0000000 +02:00;PT3M15S;352;1;0;0;"<iframe width=""480"" h
src=""//www.youtube.com/embed/_7e5gamO3nA"" frameborder=""0"" allow=""accelerometer
clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"" allowfullscreer
</iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08-24
03:00:13.867;C70CF0616F50FC5D3279AC0DCE2132BA54FC0A11D6204352F47B7D5A15B2
08-24 03:00:13.990;2023-08-24 03:00:13.990
_adoo5PLxtE;2023-08-23;11 Notre Père;2018-11-08 10:26:34.0000000 +01:00;PT37S;86;0;0;
width=""480"" height=""270"" src=""//www.youtube.com/embed/_adoo5PLxtE"" frameborder=
allow=""accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-pi
share"" allowfullscreen></iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08
03:00:13.867;D1CE35B9704A0510CC6EC75AEBA882514C3F21CBFE0D6DEA099E0231A5E

```

```

width=""480"" height=""270"" src=""//www.youtube.com/embed/_adoo5PLxtE"" frameborder=
allow=""accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-pi
share"" allowfullscreen></iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08
03:00:13.867;D1CE35B9704A0510CC6EC75AEBA882514C3F21CBFE0D6DEA099E0231A5E
08-24 03:00:13.990;2023-08-24 03:00:13.990
_as1hv2yjc;2023-08-23;Online-Gottesdienst / Culte en ligne 14.06.2020 RFST Timeout;202
00:26:36.0000000 +02:00;PT51M10S;1926;24;4;0;"<iframe width=""480"" height=""270""
src=""//www.youtube.com/embed/_as1hv2yjc"" frameborder=""0"" allow=""accelerometer; at
clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"" allowfullscreer
</iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08-24
03:00:13.867;B3C7D17661A829497F3C551EAC8CB05FD3BBE4A2AE37F4AEA342A8FDB1K
08-24 03:00:13.990;2023-08-24 03:00:13.990
_BDiejswf9Y;2023-08-23;Heilsarmee Weihnachtsvideo 2022 TVC 15sec FR 1080p25 R128 M/
4444;2022-11-25 20:36:37.0000000 +01:00;PT15S;1;0;0;0;"<iframe width=""480"" height=""27
src=""//www.youtube.com/embed/_BDiejswf9Y"" frameborder=""0"" allow=""accelerometer; a
clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"" allowfullscreer
</iframe>";0;con_ConnectServer.YouTubeAnalytics1.Videos;2023-08-24
03:00:13.867;61771CAB597BCB1FBC0C7C97E7A5C0A4155F5FC1CC44F6BB4B753DDC89B
08-24 03:00:13.990;2023-08-24 03:00:13.990
=====

```

Please create a sample dataset in csv file format using the example already provided with ran generated data. Give the Youtube videos a title that fits a product tutorial for technical medic:



Sure, here's a small sample dataset based on the structure you provided, but with the theme being "a product tutorial for technical medical products." The data is in CSV format:

CSV Copy code

```

Id,As_Of_Date,Title,PublishedAt,Duration,ViewCount,LikeCount,DislikeCount,Co
_4jKJzH432a,2023-08-23,"Tutorial: Calibration of Cardiac Monitor XZ100",2019
_3hYKwS765b,2023-08-23,"Guide: Setting Up Infusion Pump Model Y",2018-02-27
_9nBvCx890k,2023-08-23,"How to Use Respiratory Ventilator VQ500",2017-09-12
_1pLQeT432x,2023-08-23,"Introduction to MRI Machine AlphaZ",2016-04-19 15:20
_8zXvDr0981,2023-08-23,"Maintaining & Cleaning Hemodialysis Machine HDM-9",2

```

This CSV content uses random data and medical tutorial titles. Please note that the video links, hash values, and other details are made up. If you have specific requirements or constraints, let me know and I can adjust the data accordingly!

ChatGPT hat passende Testdaten erzeugt

This PC > Local Disk (C:) > temp

Search temp

Name	Date modified	Type	Size
gpt_youtube_data.csv	04/08/2023 13:41	Microsoft Excel Com...	2 KB
gpt_youtube_data_2.csv	04/08/2023 14:12	Microsoft Excel Com...	5 KB

gpt_youtube_data.csv - Excel

Search

File Home Insert Draw Page Layout Formulas Data Review View Automate Help

Clipboard: Paste, Cut, Copy, Format Painter

Font: Calibri, 11, Bold, Italic, Underline, Color, Background Color

Alignment: Left, Center, Right, Indent, Wrap Text, Merge & Center

Number: General, Percentage, Currency, Date, Time, Text, Fraction, Scientific

Conditional Formatting: Normal, Bad, Good, Neutral

Format as Table

Insert

A1: Id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Id	As_Of_Date	Title	Published	Duration	ViewCount	LikeCount	DislikeCount	CommentCount	EmbedHTML	psa_load	psa_recorder	psa_load	hash_value	dss_update	dss_create_time	
2	pvn0qP8tLO	20/06/2023	Defibrillator Operation Tutorial	2019-10-1	PT3M12S	2345	230	5	12	<iframe sr	0	con_Conn	00:19.1	E58C2A04	00:19.2	00:19.2	
3	Pwr1HN_4JFc	20/06/2023	ECG Machine How-to Guide	2010-11-2	PT2M2S	1200	115	3	8	<iframe sr	0	con_Conn	00:19.1	FFEC2DF0	00:19.2	00:19.2	
4	PX1nkHktJw	20/06/2023	Ultrasound Machine Setup Tutorial	2019-07-1	PT16S	789	68	2	4	<iframe sr	0	con_Conn	00:19.1	7A7DA0AF	00:19.2	00:19.2	
5	pY6ToHSXJA	20/06/2023	How to Use a Ventilator Machine	2022-11-2	PT2M3S	6543	620	10	15	<iframe sr	0	con_Conn	00:19.1	6C65A2D9	00:19.2	00:19.2	
6	Pyv7yD3png	20/06/2023	In-depth CPAP Machine Tutorial	2011-06-0	PT2M49S	5098	487	8	13	<iframe sr	0	con_Conn	00:19.1	DBED49D	00:19.2	00:19.2	
7																	

Die Daten lassen sich mit M-Query in Power BI importieren

Digital Dashboard_cdata_Youtube - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply Close New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Query Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Text Analytics Vision Azure Machine Learning AI Insights

Queries [6]

- DirectMarketing [3]
 - StartYear (2018)
 - EndYear (2023)
 - YT Videos Testdata
- Other Queries [3]

```
= Csv.Document(File.Contents("C:\Temp\gpt_youtube_data.csv"),[Delimiter=";", Columns=16, Encoding=1252, QuoteStyle=QuoteStyle.None])
```

AB_C Column1	AB_C Column2	AB_C Column3	AB_C Column4	AB_C Column5	AB_C Column6	AB_C Column7	AB_C Column8
1 Id	As_Of_Date	Title	PublishedAt	Duration	ViewCount	LikeCount	DislikeCount
2 pvn0qP8tL0	2023-06-20	Defibrillator Operation Tutorial	2019-10-18 15:56:30.0000000 +02:00	PT3M12S	2345	230	5
3 Pwr1HN_4JFc	2023-06-20	ECG Machine How-to Guide	2010-11-24 16:37:01.0000000 +01:00	PT2M2S	1200	115	3
4 PX1nkHktJw	2023-06-20	Ultrasound Machine Setup Tutorial	2019-07-12 13:22:44.0000000 +02:00	PT16S	789	68	2
5 pY6ToHSXJA	2023-06-20	How to Use a Ventilator Machine	2022-11-26 20:00:17.0000000 +01:00	PT2M3S	6543	620	10
6 Pyv7yD3png	2023-06-20	In-depth CPAP Machine Tutorial	2011-06-03 11:26:41.0000000 +02:00	PT2M49S	5098	487	8

Der Power BI Report ist mit den neuen Daten noch immer lauffähig.
Die Daten haben die korrekte Struktur behalten.

Digital Dashboard_cdata_Youtube - Power BI Desktop

File Home Insert Modeling View Optimize Help External tools

Paste Cut Copy Format painter Clipboard Get data Excel OneLake data SQL Server Enter Data Datasource Recent sources Transform Refresh data New visual Text box More visuals New measure Quick measure Sensitivity Publish Share

YT alle Videos

5 15975
Anzahl Videos Views

Aktionen mit Spenden > 0

	Views	Likes	Dislikes	Comments
All	15,975	1,520	28	52
How to Use a Ventilator Machine	6,543	620	10	15
In-depth CPAP Machine Tutorial	5,098	487	8	13
Defibrillator Operation Tutorial	2,345	230	5	12
ECG Machine How-to Guide	1,200	115	3	8
Ultrasound Machine Setup Tutorial	789	68	2	4

Title	Video URL	PublishedAt	Duration
Defibrillator Operation Tutorial	🔗	10/18/2019	PT3M12S
ECG Machine How-to Guide	🔗	11/24/2010	PT2M2S
How to Use a Ventilator Machine	🔗	11/26/2022	PT2M3S
In-depth CPAP Machine Tutorial	🔗	6/3/2011	PT2M49S
Ultrasound Machine Setup Tutorial	🔗	7/12/2019	PT16S



Beispiel 3 - Datentransformationen

- Wie kann Chat GPT Ihnen beim Aufbau Ihrer Datenanalyseplattform helfen?

 - Eine Datenanalyseplattform benötigt Transformationslogik
 - Logik selbst programmieren
 - ein DWH-Automatisierungstool verwenden
 - generische Logik wiederverwenden
- oder?
- Chat GPT verwenden!



Beispiel 3– Chat GPT GUI

- › Laden von PatientData in eine Dimension
 - › Erstellen der Dimensionstabelle

Could you generate an SQL server create table script for a dimension with the following columns. The first column is called dim_patient_sk which is a surrogate key maintained by SQL Server.

The table name should be dwh.dim_patient

“

Patient_ID: integer starting at 1 counting up,

Citizenship: two letter ISO country code CH or FR or DE or IT,

First_Name: first name matching citizenship,

Last_Name: last name matching citizenship,

Email: private or professional email address from random domain or email provider,

Birthdate,

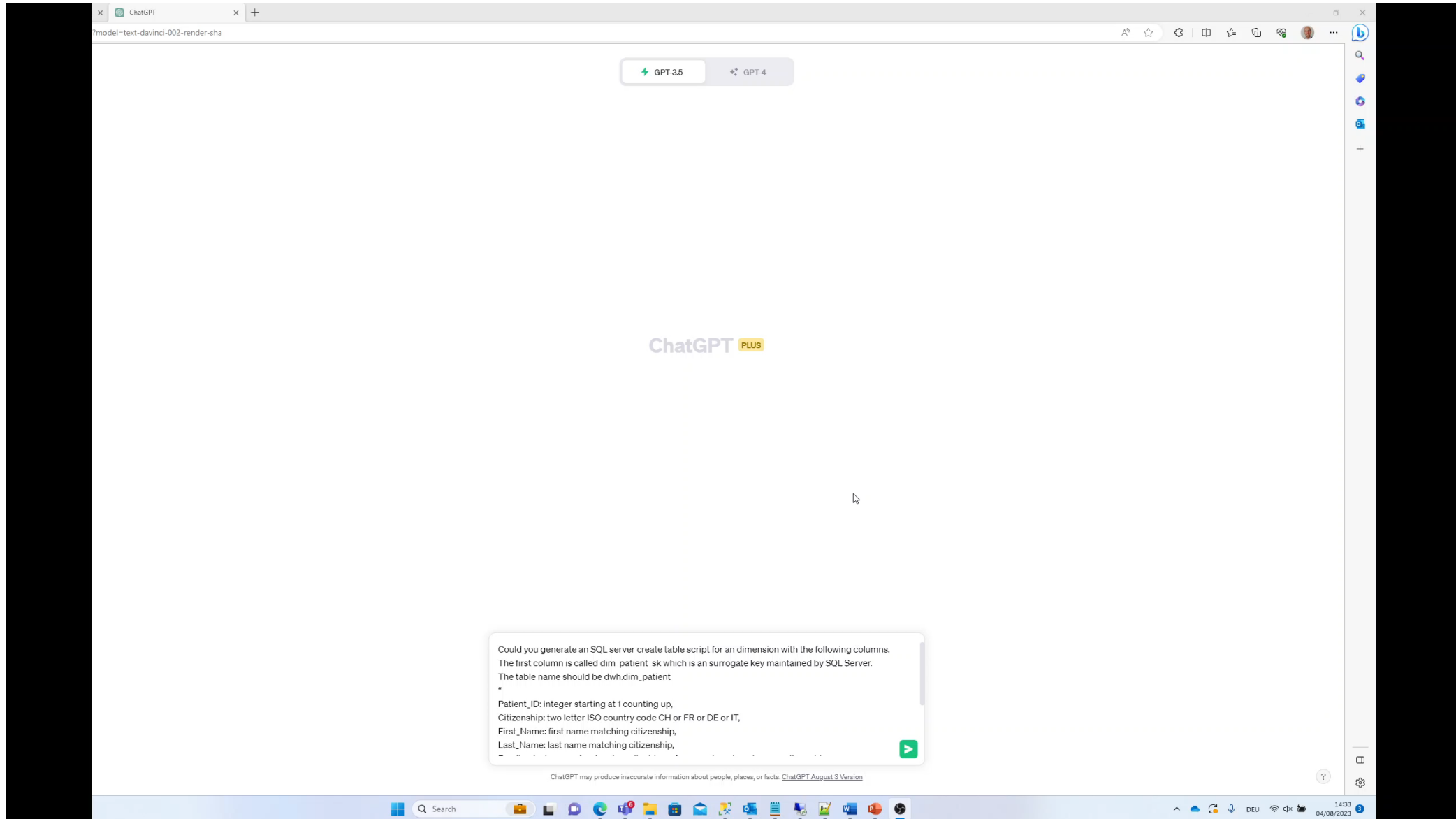
Age,

Gender_Code,

AHV_Number: Swiss social security number with a max length 20,

Main_Diagnosis: an ICD-10 code”

Teil 2 – Chat GPT GUI





Beispiel 3– Chat GPT GUI

- › Laden von PatientData in eine Dimension
 - › Erstellen einer Beladungs-Procedure

Could you generate an SQL Server stored procedure with the name `dwh.usp_scd1_dim_Patient`.

This procedure needs to merge data from the source table `dbo.PatientData` into the target table `dwh.dim_patient`.

the columns of the source table are:

"Patient_ID: integer starting at 1 counting up This is primary key,

Citizenship: two letter ISO country code CH or FR or DE or IT,

First_Name: random first name matching citizenship,

Last_Name: random last name matching citizenship,

Email: private or professional email address from random domain or email provider,

Birthdate,

Age,

Gender_Code,

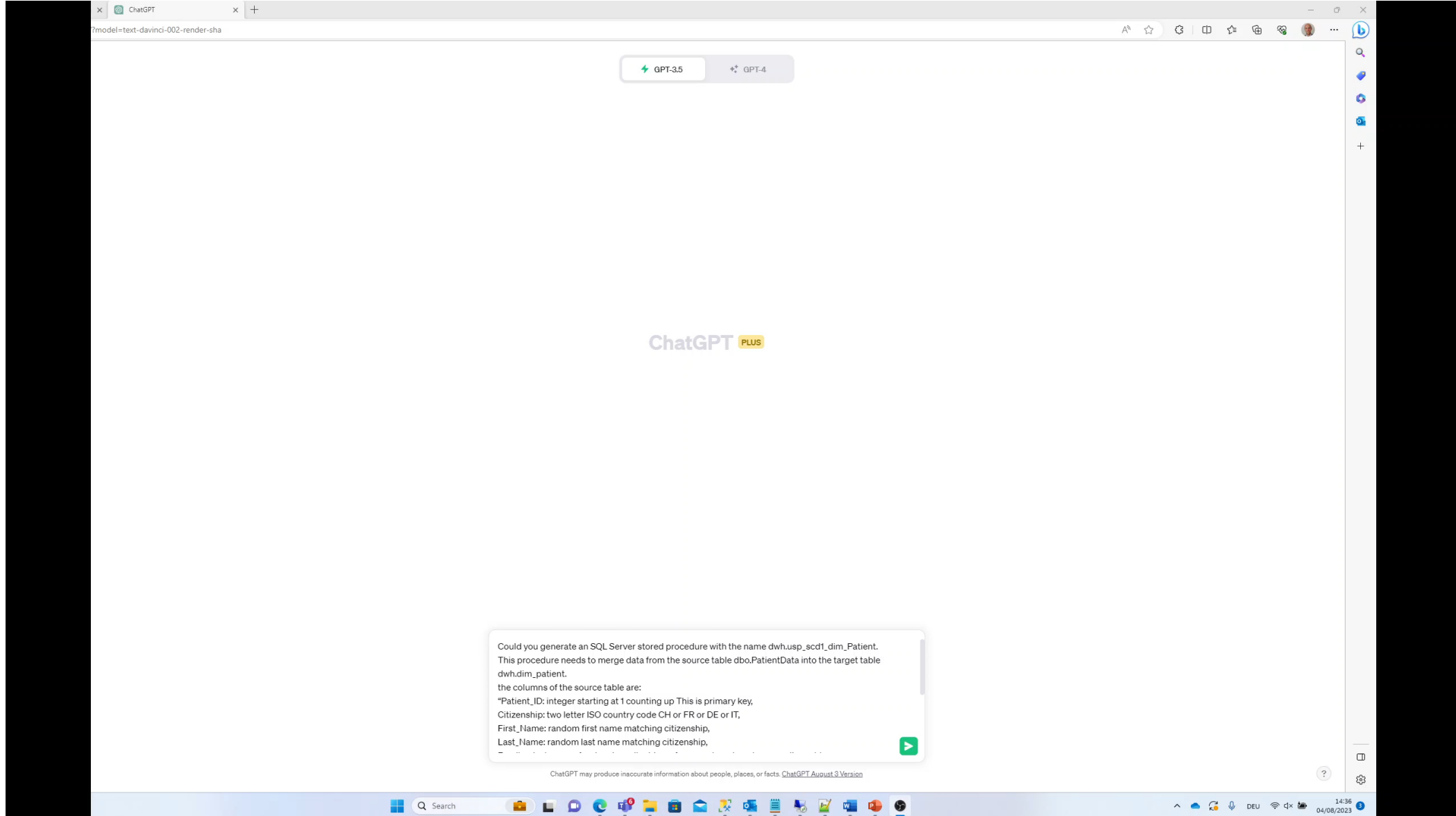
AHV_Number: random Swiss social security number in the format 756.xxxx.xxxx.xx for Swiss citizen (Citizenship CH)

or "n/a" otherwise,

Main_Diagnosis: random ICD-10 code matching the persons age and gender."

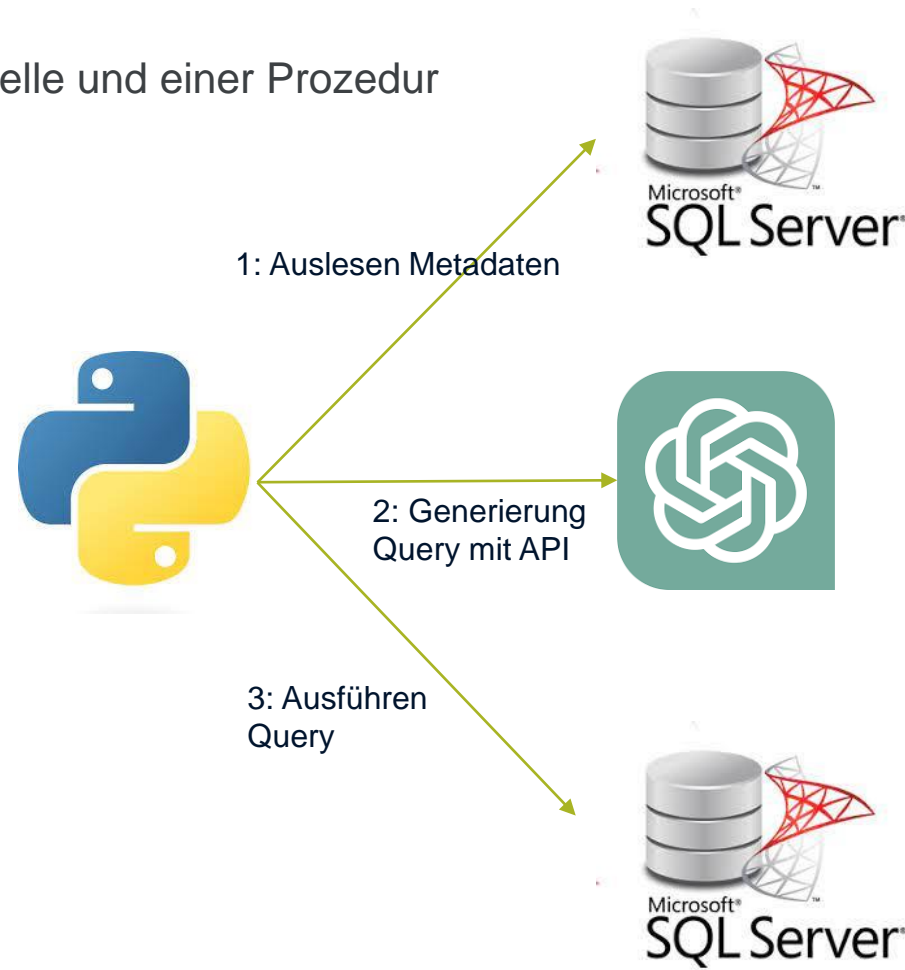
The procedure needs to merge the data from the source into the target using a slowly changing dimension type approach according to Kimball.

Teil 2 – Chat GPT GUI



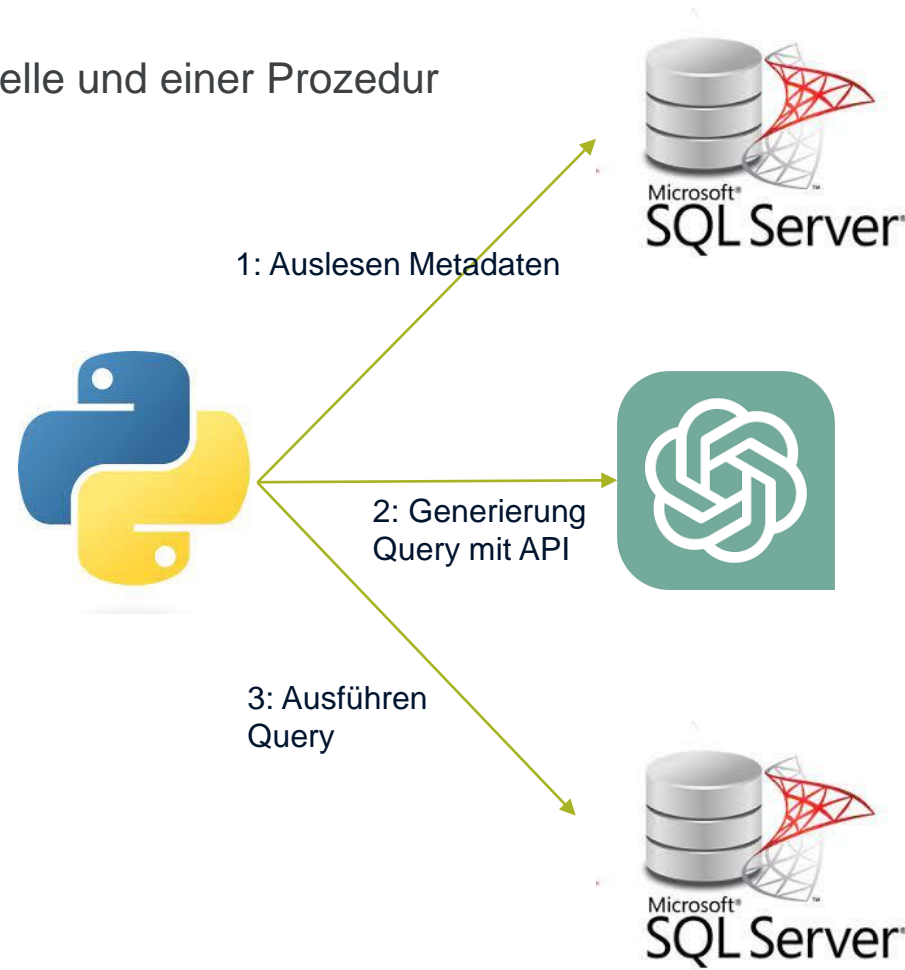
Teil 2 – Chat GPT API

- › Laden von PatientData mittels API
 - › Automatisches Erstellen einer Tabelle und einer Prozedur



Teil 2 – Chat GPT API

- › Laden von PatientData mittels API
 - › Automatisches Erstellen einer Tabelle und einer Prozedur

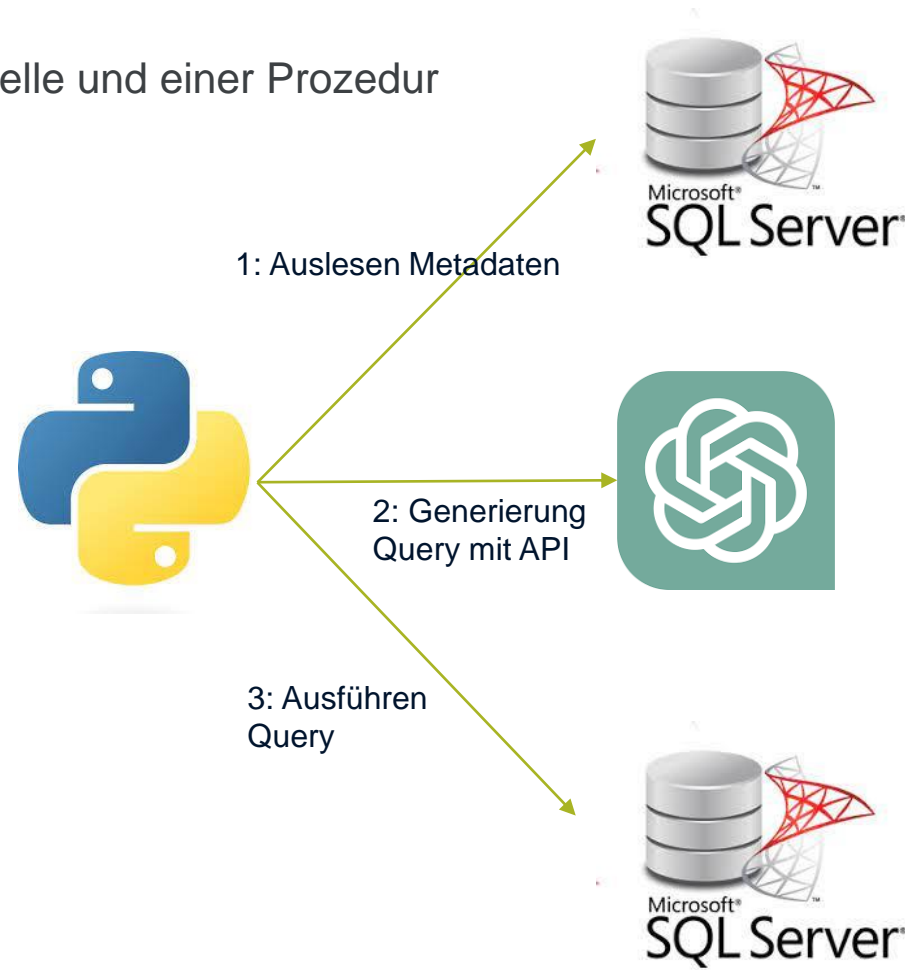


Python Function:
`retrieve_table_info(tablename)`

returns list of column names

Teil 2 – Chat GPT API

- › Laden von PatientData mittels API
 - › Automatisches Erstellen einer Tabelle und einer Prozedur

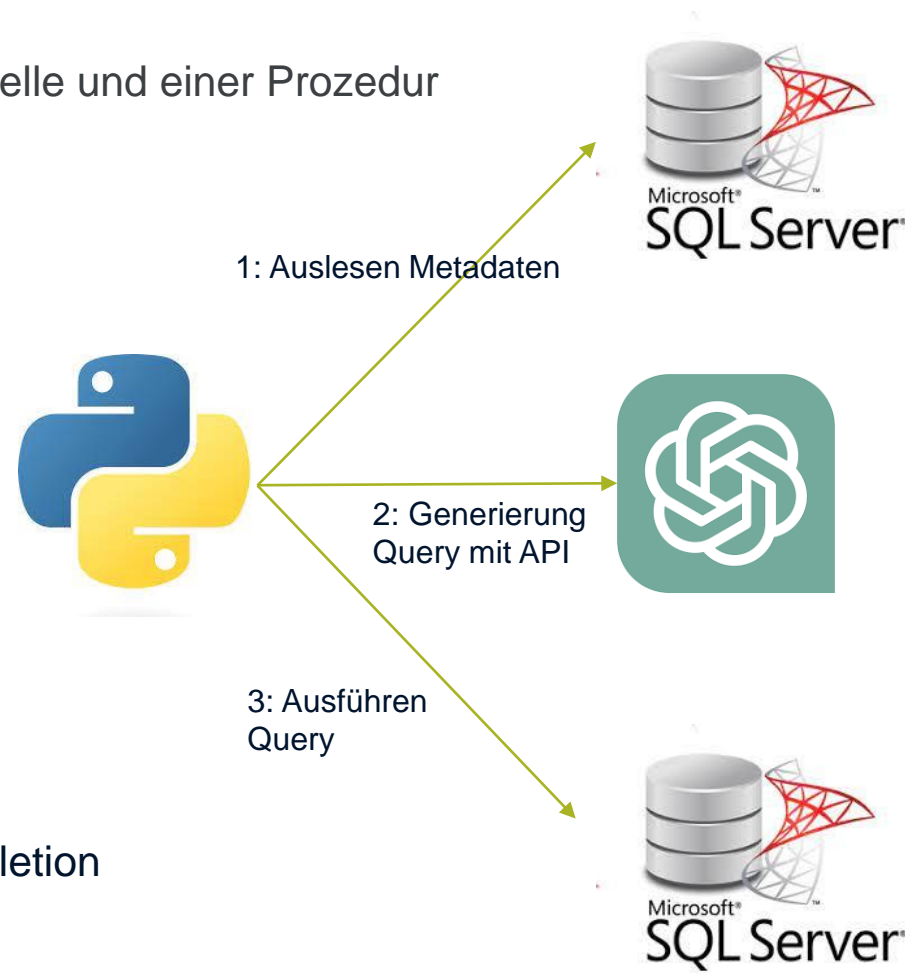


Python Function:
ask_chatgpt(prompt)

returns the answer of ChatGPT

Teil 2 – Chat GPT API

- › Laden von PatientData mittels API
 - › Automatisches Erstellen einer Tabelle und einer Prozedur



Python Function:
`execute_query(query)`

returns an message about the completion



Teil 2 – Chat GPT API – Tabelle anlegen

The screenshot shows a Databricks workspace interface. At the top, there's a header bar with 'Saved', a search bar, and a 'Fabric Trial: 59 days left' notification. Below this is a toolbar with 'Run', 'Data', and 'View' tabs, and buttons for 'Run all', 'Stop session', 'Language' (set to 'PySpark (Python)'), and 'Open in VS Code'. A warning banner states: 'Synapse notebooks and Spark job definitions are in Preview. Other users in your organization may have access to this workspace. Do not use these items unless you trust all other users who...'. The main area displays a code cell with the following Python code:

```
1 results = retrieve_table_info('PatientData')
2
3 prompt = """
4 I want to create an table in SQL server for an slowly changing dimension type 1.
5 The name of the table should be dwh.dim_Patient.
6 the Columns should be
7 - Dim_customer_sk = an surrogate key that is automatically generated.
8 """
9
10 prompt2 = """
11 - audit_insert_date = the datetime when this row has been inserted
12 - audit_update_date = the datetime when this row has been inserted
13
14 Could you generate an DDL create Table statement?
15 """
16 combined_string = prompt + str(results) + prompt2
17
18 print(combined_string)
19
20 generatedQuery = ask_chatgpt(combined_string)
21 print(generatedQuery)
22
23 result = execute_query(generatedQuery)
```

Below the code, a status bar indicates: '-Command executed in 4 sec 853 ms by Elgar Brunott on 3:30:49 PM, 8/04/23'. The bottom of the interface shows a 'Save option: Automatic' and 'Selected 1 of 9 cells'.



Teil 2 – Chat GPT API – Prozedur generieren

1 Synapse notebooks and Spark job definitions are in Preview. Other users in your organization may have access to this workspace. Do not use these items unless you trust all other users who may h

```
1 results = retrieve_table_info('PatientData')
2
3 prompt = ""
4 Could you generate an SQL Server stored procedure with the name dwh.usp_scd1_dim_Patient.
5 This procedure needs to merge data from the source table dbo.PatientData into the target table dwh.dim_patient.
6 the columns of the source table are:
7
8 ""
9 template = ""
10
11 Use the following template for generating this procedure
12 CREATE or ALTER PROCEDURE [dwh].[usp_scd1_<table>]
13 AS
14 BEGIN
15     MERGE [dwh].[dim_<table>] AS TARGET
16     USING dbo.[<source table>] AS SOURCE
17     ON ( <compare the source ID with target ID> )
18     WHEN NOT MATCHED BY TARGET THEN
19         INSERT VALUES
20             ( <COLUMNS>
21               ,getdate() -- audit_insert_date
22               ,null -- audit_update_date
23             )
24     WHEN MATCHED
25         AND
26         <Check if one of values of the columns has changed between source. and target.>
27     THEN
28         UPDATE SET <update the target columns with the source value >
29         ,TARGET.audit_update_date = getdate()
30     ;
31 END;
32 ""
33 combined_string = prompt + columns + template
34 generatedQuery = ask_chatgpt(combined_string)
35 print(generatedQuery)
36
37 result = execute_query(generatedQuery)
```



Beispiel 3 – Zusammenfassung

- › Code generieren ist sehr einfach
- › Immer kontrollieren, niemals vertrauen
- › kleine Änderung im Prompt haben potentiell grosse Auswirkungen auf den Output
- › Beispielen generieren absolut! Automatisch nein



Beispiel 4: ETL-Code erweitern und optimieren

- › Häufig ist es ein Bedürfnis, Daten rasch und unkompliziert in Power BI zu laden. Power Query bietet einen breiten Umfang von Funktionen dazu.
- › Power Query resp. M Code ist bei komplexen Anforderungen nicht intuitiv und kann den Anwender überfordern.
- › ChatGPT unterstützt die Erstellung und/oder Optimierung von M Code auch in komplexen Szenarien.



Beispiel 4 – Code generieren (M-Query für Social Media Data)

- Die Views, Likes, Follows ect der Social Media Daten ändern sich täglich. Jeden Tag gibt es ein neues Set an Daten.
- Man kann aus diesen Zeitreihen einen Verlauf abbilden
- Meistens interessiert einem aber nur die neueste Version

Queries [6]

DirectMarketing [3]
StartYear (2018)
EndYear (2023)
YT Videos Testdata

Other Queries [3]
YT Videos
YT Videos (2)
YT Videos rewritten

Query Settings

Query Name: YT Videos (2)

Query Formula: `= Table.TransformColumnTypes(#"Renamed Columns1",{{"Views", Int64.Type}, {"Likes", Int64.Type}, {"Dislikes", Int64.Type}, {"Comments", Int64.Type}})`

Id	As_Of_Date	Title	PublishedAt	Duration	Views	Likes
1	21/06/2023	Defibrillator Operation Tutorial	18/10/2019	PT3M12S	2348	
2	21/06/2023	ECG Machine How-to Guide	24/11/2010	PT2M2S	1203	
3	21/06/2023	Ultrasound Machine Setup Tutorial	12/07/2019	PT16S	792	
4	21/06/2023	How to Use a Ventilator Machine	26/11/2022	PT2M3S	6548	
5	21/06/2023	In-depth CPAP Machine Tutorial	03/06/2011	PT2M49S	5104	
6	22/06/2023	Defibrillator Operation Tutorial	18/10/2019	PT3M12S	2351	
7	22/06/2023	ECG Machine How-to Guide	24/11/2010	PT2M2S	1206	
8	22/06/2023	Ultrasound Machine Setup Tutorial	12/07/2019	PT16S	792	
9	22/06/2023	How to Use a Ventilator Machine	26/11/2022	PT2M3S	6549	
10	22/06/2023	In-depth CPAP Machine Tutorial	03/06/2011	PT2M49S	5111	
11	23/06/2023	Defibrillator Operation Tutorial	18/10/2019	PT3M12S	2357	
12	23/06/2023	ECG Machine How-to Guide	24/11/2010	PT2M2S	1214	
13	23/06/2023	Ultrasound Machine Setup Tutorial	12/07/2019	PT16S	797	
14	23/06/2023	How to Use a Ventilator Machine	26/11/2022	PT2M3S	6552	
15	23/06/2023	In-depth CPAP Machine Tutorial	03/06/2011	PT2M49S	5116	

Query Settings

PROPERTIES

Name: YT Videos (2)

APPLIED STEPS

- Source
- Promoted Headers
- Removed Columns
- Added Custom
- Changed Type
- Added Custom1
- Changed Type1
- Removed Columns1
- Renamed Columns
- Reordered Columns
- Renamed Columns1
- Changed Type2

- Im Power Query Editor gibt es für eine Max-Funktion aber keine «Klick-Buttons»

Power Query Editor Ribbon

Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh, Advanced Editor, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Sort, Split Column, Group By, Data Type: Text, Use First Row as Headers, Replace Values, Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, Azure Machine Learning

Group By, Use First Row as Headers, Count Rows, Transpose, Reverse Rows, Detect Data Type, Fill, Pivot Column, Convert to List, Split Column, Format, Parse, Merge Columns, Extract, Trigonometry, Rounding, Information, Date, Time, Duration, Date & Time Column

Column From Examples, Custom Column, Invoke Custom Function, Conditional Column, Index Column, Duplicate Column, From Text, From Number, From Date & Time, Text Analytics, Vision, Azure Machine Learning, AI Insights



M-Query Anpassung für ein „Max Date“

Chat GPT Prompt:

I have created the following M-Query code:

""""

Copy M-Query query here

""""

Now I want to have a new column named "current" which will show me the newest row with a 1 and all others with a 0. I want the new column to be based on the [As_Of_Date] column, and the logic is: if [As_Of_Date] = max [As_Of_Date] then 1 else 0.

Please integrate this function into the M-Query and explain to me what you are doing.

I have created the following M-Query code:

```
""""
let
    Source = Sql.Database(" ", "DWH", [Query="SELECT [Id]#(If) ,[As_Of_Date]#
(If) ,[Title]#(If) ,[PublishedAt]#(If) ,[Duration]#(If) ,[ViewCount]#(If) ,[LikeCount]#
(If) ,[DislikeCount]#(If) ,[CommentCount]#(If) ,[EmbedHtml]#(If) ,[psa_load_id]#(If)
,[psa_record_source]#(If) ,[psa_load_date]#(If) ,[hash_value]#(If) ,[dss_update_time]#
(If) ,[dss_create_time]#(If) FROM [ ].[yot_out].[psa_YOT_Videos_cur]")),
    #"Removed Columns" = Table.RemoveColumns(Source,{"psa_load_id",
"psa_record_source", "psa_load_date", "hash_value", "dss_update_time", "dss_create_time"}),
    #"Added Custom" = Table.AddColumn(#"Removed Columns", "Video URL", each
"https://www.youtube.com/watch?v=" & [Id]),
    MaxDate = List.Max(#"Added Custom"[As_Of_Date]),
    #"Added Current" = Table.AddColumn(#"Added Custom", "current", each if [As_Of_Date] =
MaxDate then 1 else 0),
    #"Filtered Rows" = Table.SelectRows(#"Added Current", each ([current] = 1)),
    #"Changed Type" = Table.TransformColumnTypes(#"Filtered Rows",{{"As_Of_Date", type
date}, {"PublishedAt", type date}}),
    #"Removed Columns1" = Table.RemoveColumns(#"Changed Type",{"current"}),
    #"Renamed Columns" = Table.RenameColumns(#"Removed Columns1",{{"ViewCount",
"Views"}, {"LikeCount", "Likes"}, {"DislikeCount", "Dislikes"}, {"CommentCount",
"Comments"}})
in
    #"Renamed Columns"
""""
```

Now I want to have a new column named "current" which will show me the newest row with a 1 and all others with a 0. I want the new column to be based on the [As_Of_Date] column, and the logic is: if [As_Of_Date] = max [As_Of_Date] then 1 else 0. Please integrate this function into the M-Query and explain to me what you are doing.

It looks like you've already started implementing the logic for the "current" column based on the `[As_Of_Date]` in your M-Query code. Your code fetches data from a SQL database, removes unnecessary columns, adds a video URL, identifies the maximum `[As_Of_Date]`, filters for rows with this maximum date, and makes some other adjustments.

Let's tweak and refine your code to cleanly integrate the "current" column:

```
m
Copy code

let
    Source = Sql.Database(" ", "DWH", [Query="SELECT ... FROM "
    #"Removed Columns" = Table.RemoveColumns(Source, {"psa_load_id", "psa_re
    #"Added Custom" = Table.AddColumn(#"Removed Columns", "Video URL", each
    MaxDate = List.Max(#"Added Custom"[As_Of_Date]),

    // Add the "current" column based on the As_Of_Date
    #"Added Current" = Table.AddColumn(#"Added Custom", "current", each if [

    #"Changed Type" = Table.TransformColumnTypes(#"Added Current", {{"As_Of_
    #"Renamed Columns" = Table.RenameColumns(#"Changed Type", {{"ViewCount",
in
    #"Renamed Columns"
```

Here's a breakdown of what has been done:

- Source:** Fetch data from SQL database.
- Removed Columns:** Remove the unnecessary columns.
- Added Custom:** Add the "Video URL" column.
- MaxDate:** Calculate the maximum date value from the `[As_Of_Date]` column. This will be used to determine the most recent rows.
- Added Current:** Add a new column named "current". For each row, if the `[As_Of_Date]` matches `MaxDate`, the "current" column gets a value of 1, otherwise 0. This step answers your primary question.
- Changed Type:** Transform the data type of columns for consistent data handling.
- Renamed Columns:** Rename a few columns to more user-friendly names.

You should now have a clear understanding of how the "current" column logic has been integrated into your M-Query.



Resultat: ChatGPT hat den M-Query Code erzeugt und dieser funktioniert

	current	Id	As_Of_Date	Title	PublishedAt	Duration	Views	Likes
1		7HDMreNhOxg	23.08.2023	Was ist Substitution ? Gespräch mit Seelsorger Christof Stoll	31.05.2021	PT3M39S	153	8
2		7k9p70qj8Tg	23.08.2023	Ein Fan von "Libéré" (Stress) / Un fan de "Libéré" (Stress)	05.12.2010	PT1M24S	510	1
3		7O2gGOvvdck	23.08.2023	Treppensteigen / Montée d'escaliers	26.05.2021	PT31S	39	0
4		7phinf					3569	13
5		7qfHK					7	0
6		7uXJK					1911	14
7		7V00B					952	20
8		7Xepq					82	3
9		7z23N					37	0
10		80koH					152	7
11		83ifEFj					5	0
12		87PKt					123	0
13		8Agk_					79	3
14		8blbxz					8	1
15		8CSD3					560	4
16		8faXP2					1378	13
17		8hW3t					0	0
18		8JNIS6					6148	18
19		8JTu09					18	0
20		8lkG76					249	6
21		8n-aQ					176	3
22		8oSXa					1224	27
23		8OUO					663	21
24		8TPKv					248	0
25		96WH					190	12
26		98p-rf					172	1
27		9E7ed					74	0
28		9Nmc					62	0
29		-9-psX					2000	24
30		9qvE1					473	7
31		9WDF					220	6
32		9zDqU					102	3
33		9zWxC					44	2
34		a1NIq					334	6
35		A1nm					640	2
36		a6uyK					84	1
37		AAD2c					8929	28
38		abb_d					34	0
39		aBbp5					382	8
40		ACD5V4zKZt8	23.08.2023	Tuormer-ipp: Ubdachlosen bestehen	17.12.2015	PT1M59S	112	2
41		Afs1_0EgGw0	23.08.2023	Alive	24.11.2010	PT5M49S	613	4

YT Videos (2)

```
1 let
2 Source = Sql.Database("...", "DW", [Query="SELECT [Id]#(lf)      ,[As_Of_Date]#(lf)      ,[Title]#(lf)      ,[PublishedAt]
3      #[lf]      ,[Duration]#(lf)      ,[ViewCount]#(lf)      ,[LikeCount]#(lf)      ,[DislikeCount]#(lf)      ,[CommentCount]#(lf)
4      ,[EmbedHtml]#(lf)      ,[psa_load_id]#(lf)      ,[psa_record_source]#(lf)      ,[psa_load_date]#(lf)      ,[hash_value]
5      #[lf]      ,[dss_update_time]#(lf)      ,[dss_create_time]#(lf) FROM [ ... ].[yot_out].[psa_YOT_Videos_cur]"])),
6 #Removed Columns" = Table.RemoveColumns(Source, {"psa_load_id", "psa_record_source", "psa_load_date", "hash_value",
7      "dss_update_time", "dss_create_time"}),
8 #Added Custom" = Table.AddColumn(#"Removed Columns", "Video URL", each "https://www.youtube.com/watch?v=" & [Id]),
9 MaxDate = List.Max(#"Added Custom"[As_Of_Date]),
10 // Add the "current" column based on the As_Of_Date
11 #Added Current" = Table.AddColumn(#"Added Custom", "current", each if [As_Of_Date] = MaxDate then 1 else 0),
12 #Changed Type" = Table.TransformColumnTypes(#"Added Current", {{"As_Of_Date", type date}, {"PublishedAt", type date}}),
13 #Renamed Columns" = Table.RenameColumns(#"Changed Type", {{"ViewCount", "Views"}, {"LikeCount", "Likes"}, {"DislikeCount",
14      "Dislikes"}, {"CommentCount", "Comments"}}),
15 #Reordered Columns" = Table.ReorderColumns(#"Renamed Columns",{"current", "Id", "As_Of_Date", "Title", "PublishedAt",
16      "Duration", "Views", "Likes", "Dislikes", "Comments", "EmbedHtml", "Video URL"}),
17 #Filtered Rows" = Table.SelectRows(#"Reordered Columns", each ([current] = 1))
18 in
19 #Filtered Rows"
```

✓ No syntax errors have been detected.

Done Cancel



M-Query Code überarbeiten

- ChatGPT Prompt:
Please simplify the following m-query by grouping similar tasks together.
Make sure, the end result of the m-query remains the same after changing the code.
Here is the M-Query Code

""""

Copy M-Query query here

""""

APPLIED STEPS

Source

Promoted Headers

Removed Columns

Added Custom

Changed Type

Added Custom1

Changed Type1

Removed Columns1

Renamed Columns

Reordered Columns

Renamed Columns1

Changed Type2

Advanced Editor

YT Videos

```
1 let
2   Source = Csv.Document(File.Contents("C:\Temp\gpt_youtub_data.csv"),[Delimiter=";", Columns=16, Encoding=1252,
3     QuoteStyle=QuoteStyle.None]),
4   #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
5   #"Removed Columns" = Table.RemoveColumns(#"Promoted Headers",{"psa_load_id", "psa_record_source", "psa_load_date", "hash_value",
6     "dss_update_time", "dss_create_time"}),
7   #"Added Custom" = Table.AddColumn(#"Removed Columns", "Video URL", each "https://www.youtube.com/watch?v=" & [Id]),
8   #"Changed Type" = Table.TransformColumnTypes(#"Added Custom",{{"As_Of_Date", type date}, {"PublishedAt", type text}}),
9   #"Added Custom1" = Table.AddColumn(#"Changed Type", "PublishedDate", each Text.Start([PublishedAt], 10)),
10  #"Changed Type1" = Table.TransformColumnTypes(#"Added Custom1",{{"PublishedDate", type date}}),
11  #"Removed Columns1" = Table.RemoveColumns(#"Changed Type1",{"PublishedAt"}),
12  #"Renamed Columns" = Table.RenameColumns(#"Removed Columns1",{"ViewCount", "Views"}, {"LikeCount", "Likes"}, {"DislikeCount",
13    "Dislikes"}, {"CommentCount", "Comments"}),
14  #"Reordered Columns" = Table.ReorderColumns(#"Renamed Columns",{"Id", "As_Of_Date", "Title", "PublishedDate", "Duration",
15    "Views", "Likes", "Dislikes", "Comments", "EmbedHtml", "Video URL"}),
16  #"Renamed Columns1" = Table.RenameColumns(#"Reordered Columns",{"PublishedDate", "PublishedAt"}),
17  #"Changed Type2" = Table.TransformColumnTypes(#"Renamed Columns1",{{"Views", Int64.Type}, {"Likes", Int64.Type}, {"Dislikes",
18    Int64.Type}, {"Comments", Int64.Type}}),
19  in
20  #"Changed Type2"
```

APPLIED STEPS

Source

Promoted Headers

Removed Columns

Added Video URL

Transformed Column Types

Added PublishedDate

Transformed PublishedDate T...

Removed PublishedAt

Renamed Columns

Reordered Columns

Final

Advanced Editor

YT Videos simplified, grouped

```
1 let
2   Source = Csv.Document(File.Contents("C:\Temp\gpt_youtub_data.csv"),[Delimiter=";", Columns=16, Encoding=1252,
3     QuoteStyle=QuoteStyle.None]),
4   #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
5   #"Removed Columns" = Table.RemoveColumns(#"Promoted Headers",{"psa_load_id", "psa_record_source", "psa_load_date", "hash_value",
6     "dss_update_time", "dss_create_time"}),
7   #"Added Video URL" = Table.AddColumn(#"Removed Columns", "Video URL", each "https://www.youtube.com/watch?v=" & [Id]),
8   #"Transformed Column Types" = Table.TransformColumnTypes(#"Added Video URL",{{"As_Of_Date", type date}, {"PublishedAt", type
9     text}}),
10  #"Added PublishedDate" = Table.AddColumn(#"Transformed Column Types", "PublishedDate", each Text.Start([PublishedAt], 10)),
11  #"Transformed PublishedDate Type" = Table.TransformColumnTypes(#"Added PublishedDate",{{"PublishedDate", type date}}),
12  #"Removed PublishedAt" = Table.RemoveColumns(#"Transformed PublishedDate Type",{"PublishedAt"}),
13  #"Renamed Columns" = Table.RenameColumns(#"Removed PublishedAt",{"ViewCount", "Views"}, {"LikeCount", "Likes"}, {"DislikeCount",
14    "Dislikes"}, {"CommentCount", "Comments"}, {"PublishedDate", "PublishedAt"}),
15  #"Reordered Columns" = Table.ReorderColumns(#"Renamed Columns",{"Id", "As_Of_Date", "Title", "PublishedAt", "Duration", "Views",
16    "Likes", "Dislikes", "Comments", "EmbedHtml", "Video URL"}),
17  Final = Table.TransformColumnTypes(#"Reordered Columns",{{"Views", Int64.Type}, {"Likes", Int64.Type}, {"Dislikes", Int64.Type},
18    {"Comments", Int64.Type}}),
19  in
20  Final
```

Beispiel 5: Code kommentieren

- › Bei komplexen Transformationen mit vielen Schritten kann es für den Entwickler schwierig werden, den Überblick über den Code zu behalten.
- › Dokumentation von Code ist mit gewissem Aufwand verbunden und wird häufig vernachlässigt.
- › ChatGPT kann bestehenden Code kommentieren

M-Query Code dokumentieren

- Zweiter Chat GPT Prompt:
- Ok, please document directly into the M-Query-Code what is done by the code, and change the step names to a description matching the tasks done in the step

APPLIED STEPS	
LoadCSV	① ✖
PromoteHeaders	① ✖
RemoveUnwantedColumns	①
GenerateVideoURL	① ✖
ConvertColumnTypes1	①
ExtractPublishDate	① ✖
ConvertPublishedDateType	①
DropOriginalPublishedAt	①
RenameForClarity	①
ReorderForLayout	①
✖ FinalizeDataTypes	①

Id	As_Of_Date	Title	PublishedAt	Duration	Views	Likes	Dislikes
1 pvn0qP8tL0	20.06.2023	Defibrillator Operation Tutorial	18.10.2019	PT3M12S		2345	230
2 Pwr1HN_4JFc	20.06.2023	ECG Machine How-to Guide	24.11.2010	PT2M2S		1200	115
3 PK1nkHktJw	20.06.2023	Ultrasound Machine Setup Tutorial	12.07.2019	PT16S		789	68
4 pY6ToHSKJA	20.06.2023	How to Use a Ventilator Machine	26.11.2022	PT2M3S		6543	620
5 Pvv7yD3png	20.06.2023	In-depth CPAP Machine Tutorial	03.06.2011	PT2M49S		5098	487

YT Videos documented		Display Options
<pre>1 let 2 // Load the CSV file into a table 3 LoadCSV = Csv.Document(File.Contents("C:\Temp\gpt_youtub_data.csv"),[Delimiter=";", Columns=16, Encoding=1252, QuoteStyle=QuoteStyle.None]), 4 5 // Promote the first row to column headers 6 PromoteHeaders = Table.PromoteHeaders(LoadCSV, [PromoteAllScalars=true]), 7 8 // Remove unnecessary columns 9 RemoveUnwantedColumns = Table.RemoveColumns(PromoteHeaders, {"psa_load_id", "psa_record_source", "psa_load_date", "hash_value", "dss_update_time", 10 "dss_create_time"}), 11 12 // Add a column to generate YouTube video URLs using the Id 13 GenerateVideoURL = Table.AddColumn(RemoveUnwantedColumns, "Video URL", each "https://www.youtube.com/watch?v=" & [Id]), 14 15 // Change the data types of the As_Of_Date and PublishedAt columns 16 ConvertColumnTypes1 = Table.TransformColumnTypes(GenerateVideoURL, {"As_Of_Date", type date}, {"PublishedAt", type text}), 17 18 // Extract the date from the PublishedAt column and create a new PublishedDate column 19 ExtractPublishDate = Table.AddColumn(ConvertColumnTypes1, "PublishedDate", each Text.Start([PublishedAt], 10)), 20 21 // Change the data type of the PublishedDate column to date 22 ConvertPublishedDateType = Table.TransformColumnTypes(ExtractPublishDate, {"PublishedDate", type date}), 23 24 // Remove the original PublishedAt column 25 DropOriginalPublishedAt = Table.RemoveColumns(ConvertPublishedDateType, {"PublishedAt"}), 26 27 // Rename several columns for clarity 28 RenameForClarity = Table.RenameColumns(DropOriginalPublishedAt, {"ViewCount", "Views"}, {"LikeCount", "Likes"}, {"DislikeCount", "Dislikes"}, 29 {"CommentCount", "Comments"}, {"PublishedDate", "PublishedAt"}), 30 31 // Reorder columns for a better layout 32 ReorderForLayout = Table.ReorderColumns(RenameForClarity, {"Id", "As_Of_Date", "Title", "PublishedAt", "Duration", "Views", "Likes", "Dislikes", "Comments", 33 "EmbedHtml", "Video URL"}), 34 35 // Convert columns to the appropriate data types 36 FinalizeDataTypes = Table.TransformColumnTypes(ReorderForLayout, {"Views", Int64.Type}, {"Likes", Int64.Type}, {"Dislikes", Int64.Type}, {"Comments", 37 Int64.Type}), 38 39 in 40 FinalizeDataTypes</pre>		



Data Privacy – was geschieht mit den Daten?

- › OpenAI & Co sammeln die Daten, die als «User Content» u.a. in Form von Prompts eingegeben werden für eine Reihe von Zwecken
- › die Infrastruktur von OpenAI steht in den USA und unterliegt den dortigen Reglementierungen
- › auf keinen Fall sollten sensitive oder vertrauliche Daten in irgendeiner Form ohne vorgängige Abklärung der datenschutzrechtlichen Situation verarbeitet werden
- › OpenAI und auch Microsoft bieten Business Services an, die striktere Anforderungen an den Datenschutz erfüllen, z.B.
 - › [ChatGPT Enterprise \(openai.com\)](https://openai.com)
 - › [Azure OpenAI Service – Large Language Models for Generative AI \(microsoft.com\)](https://microsoft.com)
NB: «Switzerland North» kann als Rechenzentrum gewählt werden!



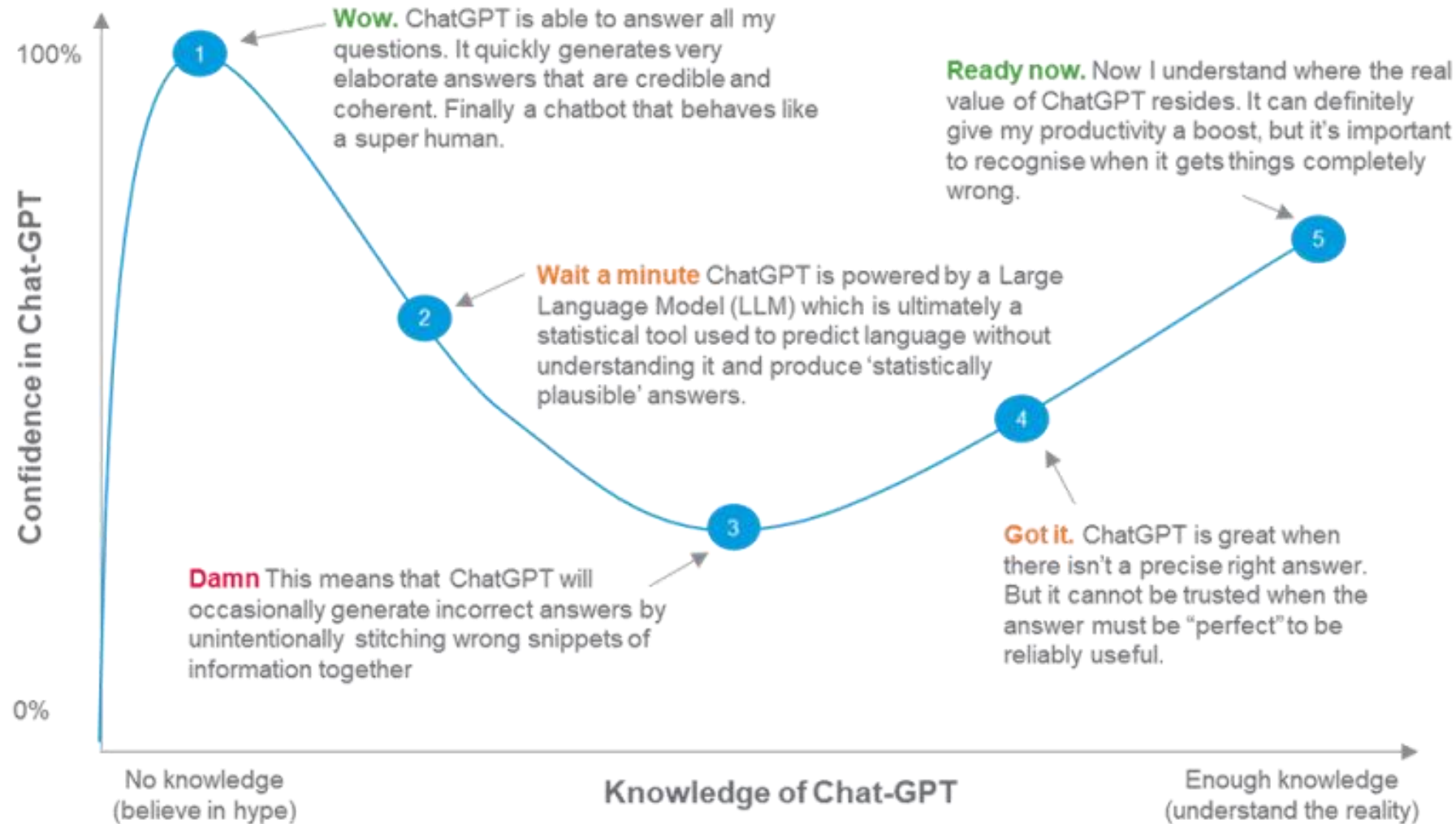
Fazit

- › «Prompt-Engineering» – es spielt eine grosse Rolle, wie man Prompts formuliert und strukturiert
- › Determinismus – nein ;)
- › faktische Korrektheit – verlass Dich nicht darauf ;)

- › Limiten
 - › Prompt-Länge
 - › Komplexität – unterbrechen in «verdaubare» Stücke

- › **ChatGPT hat ein enormes Potential!**

Adaptation von ChatGPT in die eigene Arbeit

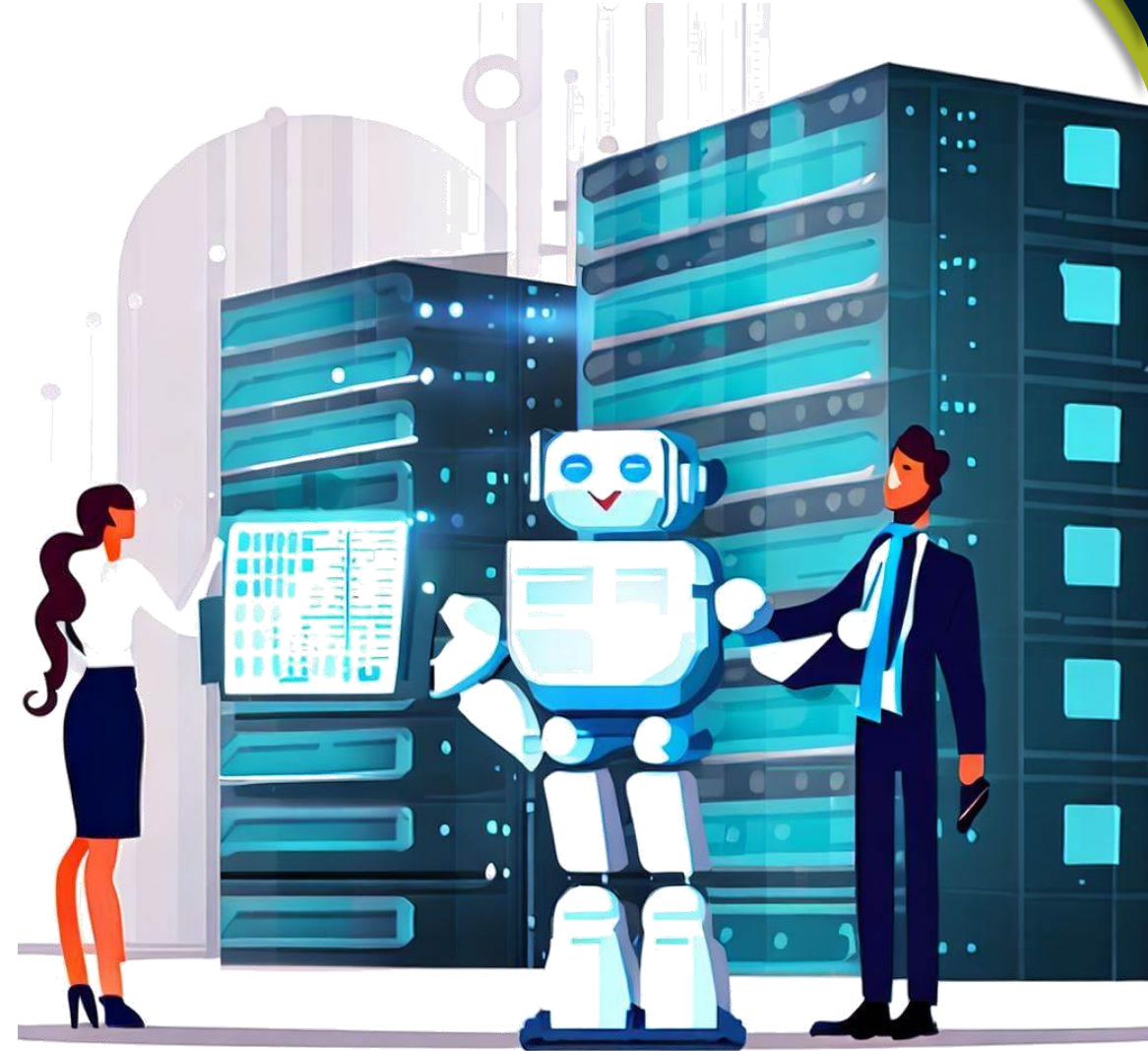


adaptiert von <https://www.rsmuk.com/insights/industry-insights/generative-ai-revolution-navigating-the-hype-cycle>





Ausblick

- › Unser neu geschaffenes «AI-Board» koordiniert die Nutzung und Weiterentwicklung von AI innerhalb IT-Logix.
- › Eine der ersten Aufgabenstellungen für das AI-Board ist es, die rechtlichen Rahmenbedingungen für den Einsatz von ChatGPT & Co. in Kundenprojekten zu definieren.
- › Dafür prüfen wir u.a. ChatGPT Enterprise sowie die Nutzung einer eigenen ChatGPT-Instanz in Azure.
- › Aktuell sind wir daran, unser Beratungsangebot für ChatGPT & Co. zu entwickeln – dazu suchen wir interessierte Unternehmen, um in der Diskussion mit diesen Unternehmen ein optimales Angebot zu schnüren. Melden Sie sich gerne bei uns: contact@it-logix.ch oder direkt nach dem Vortrag!





Bewerten Sie diese Session jetzt in der virtuellen Plattform

- 1 Loggen Sie sich auf der virtuellen Plattform mit **Ihrem persönlichen Link** ein (aus Bestätigungs-Mail MAKE BI)
- 2 Kicken Sie in der Lobby auf der **Session B4** auf den Button „**Teilnehmen**“

- 3 Kicken Sie in der **Session B4** auf der rechten Seite in der Navigation auf „**Umfragen**“
MOBILE: Klicken Sie der Navigation unten zuerst auf „**Interaktion**“, danach auf das **Umfragen-Symbol**

- 4 Beantworten Sie die vier Fragen, indem Sie pro Frage auf das zutreffende Feld klicken und Ihre Auswahl anschliessend mit einem Klick auf den Button „**Abstimmen**“ bestätigen

Die von Ihnen ausgewählten Bewertungen werden weder für die anderen Teilnehmenden noch für die Referenten ersichtlich sein. Das Ergebnis der Auswertung wird nach dem Event in aggregierter und anonymisierter Form an die jeweiligen Referenten kommuniziert werden.



Twitter



LinkedIn



Xing



YouTube



Blog

