

# DATA ANALYTICS & KI IN DER REVISIONSARBEIT | USE CASES & BEST PRACTICES

DZ Bank Gruppe | Tagung der Konzernrevision

18. September 2024

**protiviti**<sup>®</sup>  
*Global Business Consulting*

# Vorstellung der Redner



Peter Grasegger

Managing Director  
Solution MD Internal Audit



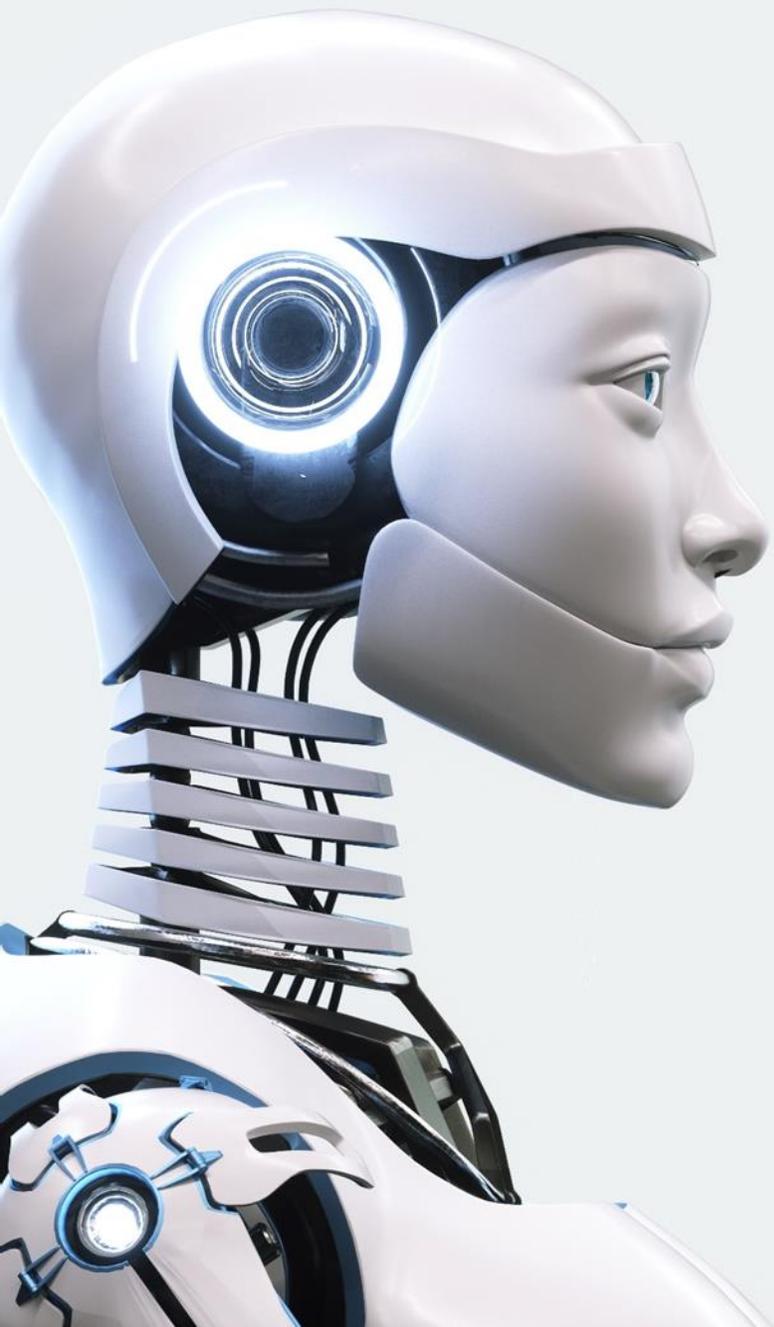
Denis Lippolt

Director  
Segment Lead Risk Analytics



Jan Zeitz

Associate Director  
Segment Risk Analytics



**01**

**Data Analytics & KI im Revisionsprozess**

**02**

**Vorstellung Use Cases:**

- 1) Vorbereitung eines Prüfprogramms
- 2) Erstellung einer Prozessdokumentation aus virtuellen Interviews
- 3) Risikobasierte Sample-Generierung mit GPT
- 4) Entwicklung eines Berichtsentwurfs aus Prüfungsartefakten

**03**

**Risiken & Regulatorik**

**04**

**Best Practices**

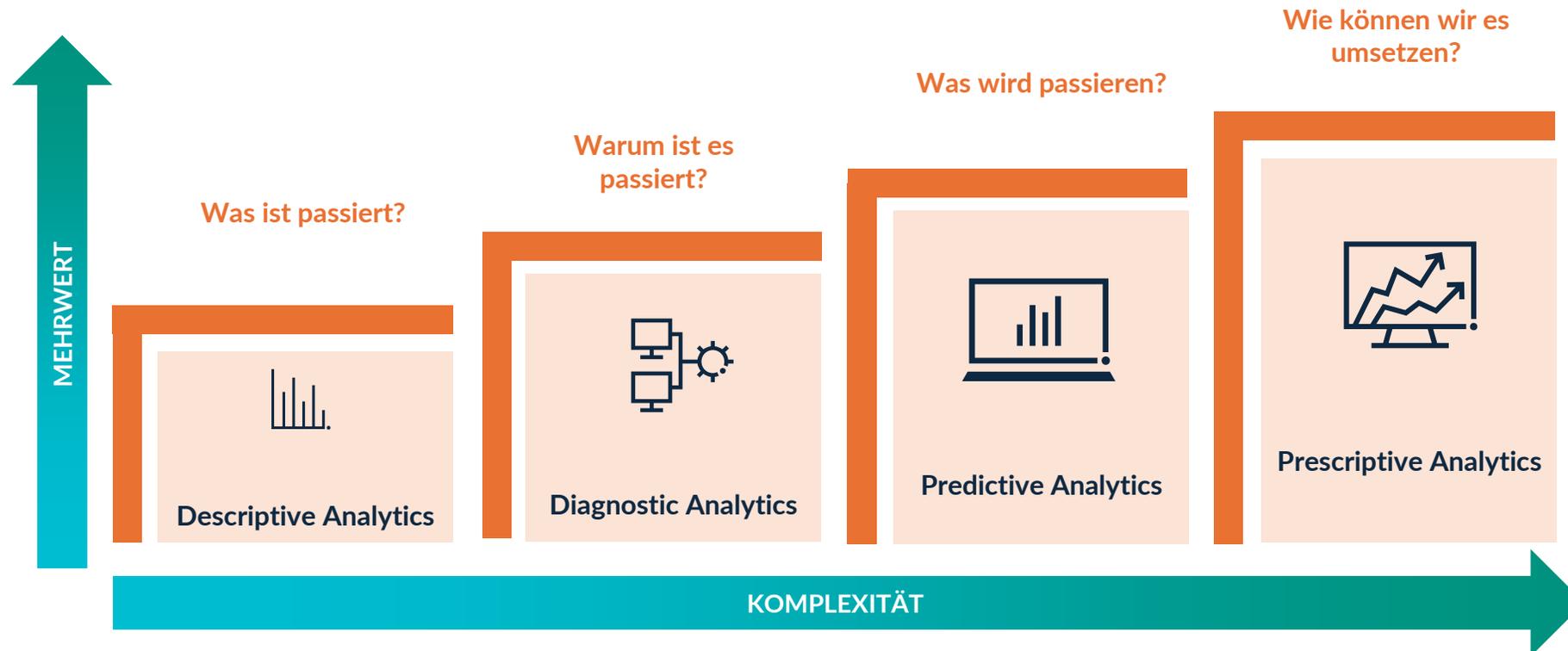
**05**

**Q&A**

# Data Analytics & KI im Revisionsprozess

# Advanced Analytics

Unter Advanced Analytics versteht man die autonome oder halbautonome Untersuchung von Daten oder Inhalten unter Verwendung ausgefeilter Techniken und Tools, die in der Regel über die der traditionellen Business Intelligence (BI) hinausgehen, um tiefere Einblicke zu gewinnen, Vorhersagen zu treffen oder Empfehlungen zu generieren.



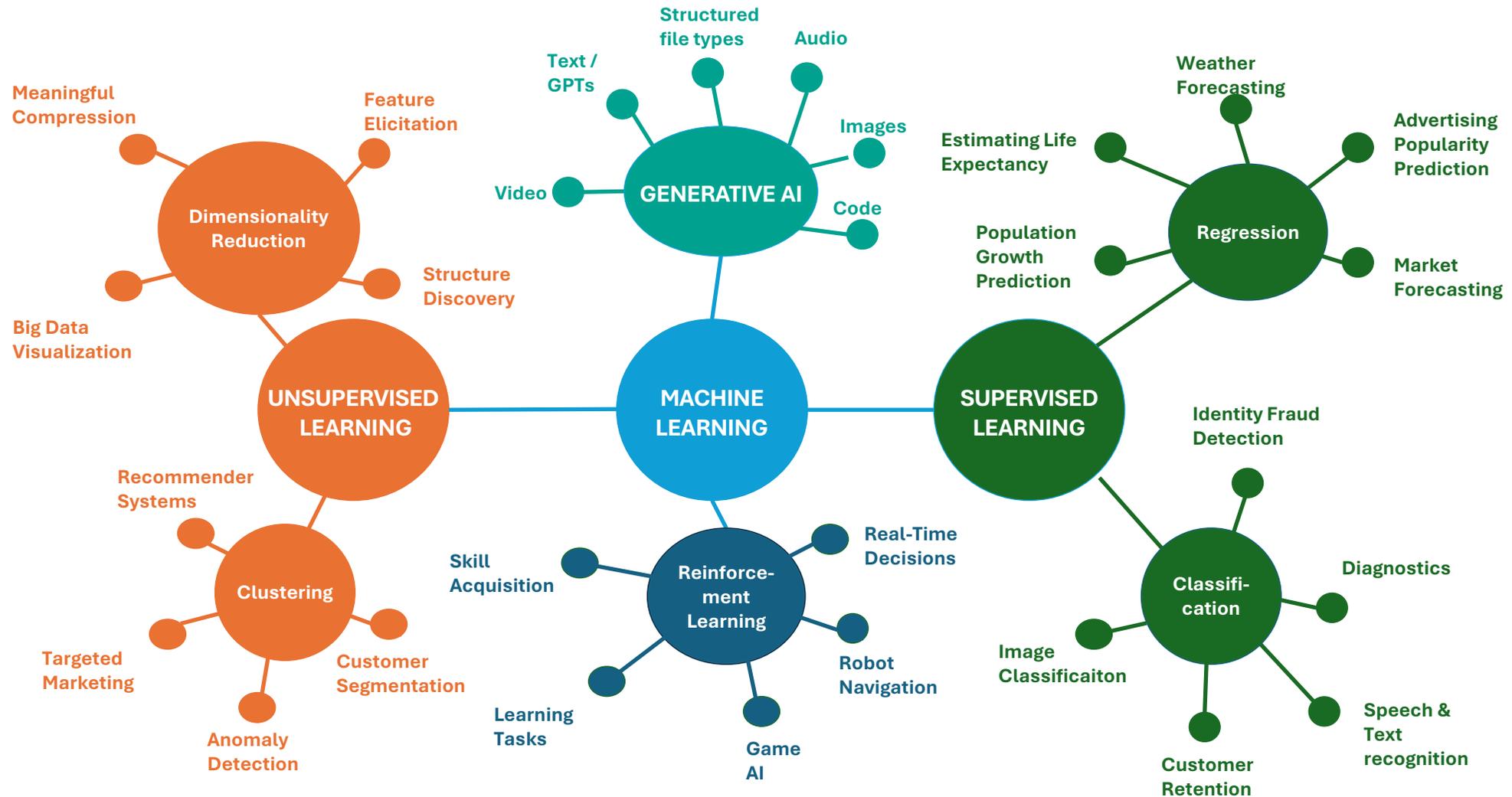
# KI Verständnis: Modelltypen

**Supervised:**  
Richtige Antwort erhalten

**Unsupervised:**  
Ähnliche Dinge gruppiert

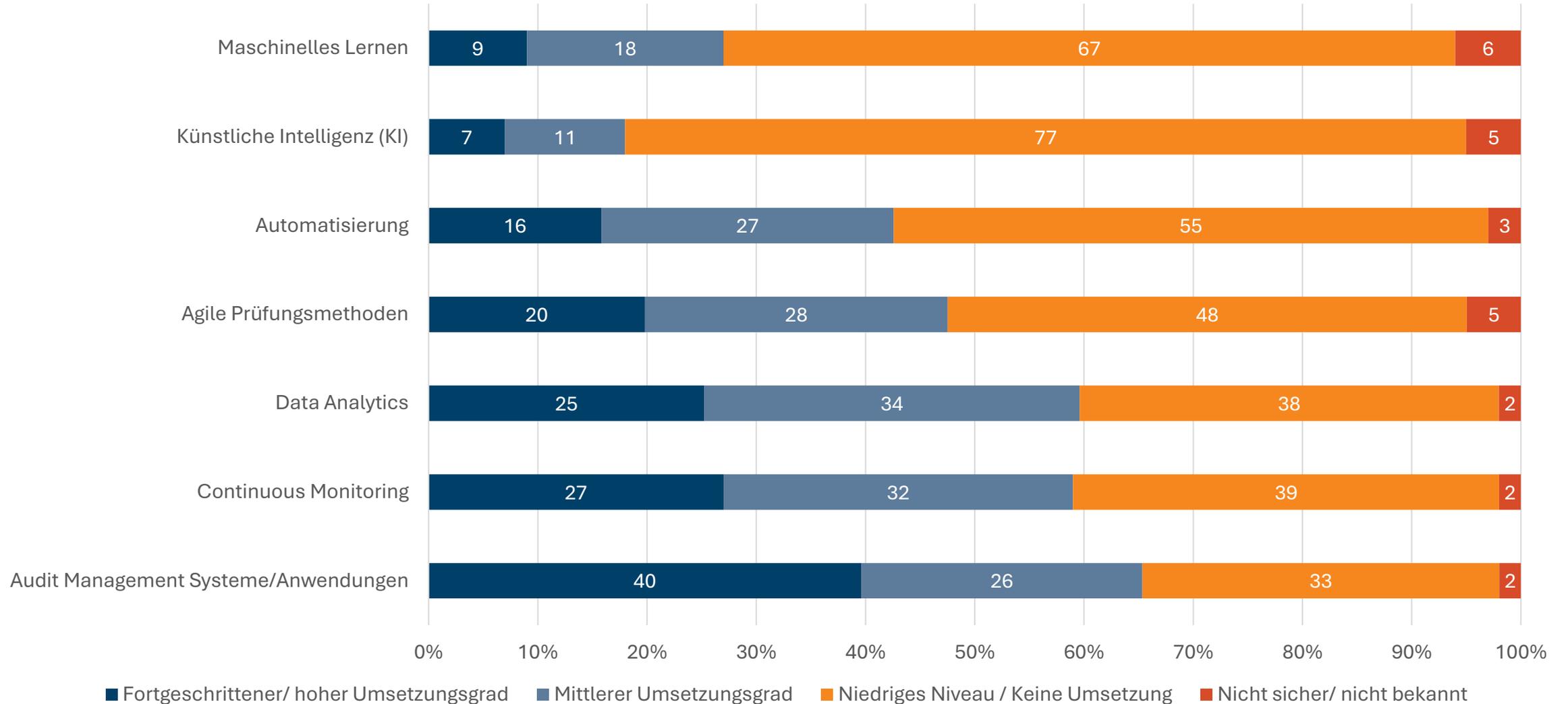
**Reinforcement:**  
learn by doing

**Generative\*:**  
Plausible Fälschungen erstellen



\*Credit to Cassie Kozyrkov, Chief Decision Scientist at Google

# IIA Internal Audit Vision 2035



# Anwendungsfälle von Data Analytics/ KI im Revisionsprozess

Planung	Durchführung	Berichterstattung	Andere Anwendungsfälle	
 Prüfungs- ankündigung	 Prozessfluss- diagramme	 Entwicklung Berichtsentwurf	 Internal Audit Charter Erstellung	 Revisionsstrategie & Roadmap
 Planungs- und Prüfungsmemorandum	 Prozess-/ Kontrollbeschreibungen	 Feststellungen & Empfehlungen	 Revisionsrichtlinien & - prozesse	 Risk Assessment
 Dokumenten Anforderungsliste	 Prozess-/ Kontrollprüfungen	 Audit Response Analyser	 Audit Analytics & Automation	 Stakeholder Kommunikation
 Prüfprogramm- entwicklung	 Prüfprogramm- abarbeitung	 Best Practices & Value Adds	 Meeting Notes Cleaner	 Research

Use Case:  
Vorbereitung eines Prüfprogramms

# ProtivitiGPT | Der intelligente KI-Assistent

The screenshot displays the ProtivitiGPT web application interface. At the top, there are navigation links for "Submit Issue", "Submit Feedback", and "KS". The main content area is titled "Chat" and "Persona". It contains a SWOT analysis in German:

- 1. Hohe Abhängigkeit von Schlüsselpersonal:** Die Leistung des Unternehmens hängt stark vom Talent einzelner Personen ab, deren Weggang den Betrieb oder die Kundenbeziehungen stören könnte.
- 2. Kostenstruktur:** Hohe Betriebskosten durch Premium-Bürostandorte und Spitzengehälter können die Preisgestaltung im Vergleich zu neuen Marktteilnehmern weniger wettbewerbsfähig machen.
- 3. Begrenzte Dienstleistungsdifferenzierung:** Eine Überabhängigkeit von bestimmten Dienstleistungsangeboten oder Sektoren könnte Risiken bei Abschwächung der Industrie bedeuten.

**Chancen**

- 1. Expansion in Schwellenmärkte:** Es besteht Raum für Expansion in Schwellenländer, die aktuell von globalen Beratungsgiganten unterversorgt sind.
- 2. Dienstleistungen zur digitalen Transformation:** Da Unternehmen zunehmend digitale Technologien übernehmen, gibt es erhebliches Wachstumspotenzial beim Angebot von Beratung zur digitalen Transformation.
- 3. Strategische Partnerschaften/Übernahmen:** Durch Allianzen oder den Kauf von Boutique-Firmen könnten Fähigkeiten und Reichweite in strategischen Bereichen oder Technologien schnell skaliert werden.

**Bedrohungen**

- 1. Marktwettbewerb:** Intensiver Wettbewerb sowohl durch etablierte Spieler als auch neue Akteure, die ähnliche Dienstleistungen zu niedrigeren Preisen anbieten.
- 2. Regulatorische Änderungen:** Änderungen in internationalen Geschäftsgesetzen könnten grenzüberschreitende Operationen komplizieren oder kostspielige Compliance-Maßnahmen erforderlich machen.
- 3. Wirtschaftsabschwünge:** Globale wirtschaftliche Instabilität könnte zu Kürzungen in Sektoren wie der Beratung führen, wo Unternehmen diese Dienste möglicherweise als nicht wesentlich ansehen.

Diese SWOT-Analyse sollte der strategischen Planung bei [Insert Company] als Leitfaden dienen, um seine Stärken zu nutzen, Schwachstellen anzugehen, neue Wachstumschancen zu erfassen und sich auf potentielle Bedrohungen in der dynamischen Landschaft der Beratungsbranche vorzubereiten.

AI-generated content may be incorrect

Message ProGPT

**Suggested Prompts**

You must follow [these requirements](#) when using ProGPT. You must always review, revise, and verify ProGPT output before you use it in deliverables or thought leadership. Please note data input into ProGPT may be processed in the US.



Secure

LLM

Document Chat

# Exkurs: Prompting



**Aufgabe**



**Beispiele**



**Format**



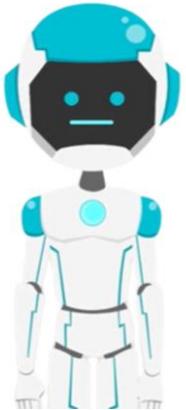
**Kontext**



**Persona**



**Tonalität**



# Use Case: Vorbereitung eines Prüfprogramms

## Kurzbeschreibung

- Hochladen der „DORA“ (Gesetzestext) als PDF in ProtivitiGPT
- Formulierung der Persona
- Eingabe eines guten Prompts, um ein risikoorientiertes Prüfprogramm zu erzeugen

## Technologie

- ProtivitiGPT
- MS Excel

## Mehrwert

- Große Zeitersparnis bei der Erstellung einer Zusammenfassung/ Risiko-Kontroll-Matrix für eine neue regulatorische Anforderung
- Möglichkeit einen ersten Überblick über eine neue regulatorische Anforderung zu bekommen

## Komplexität



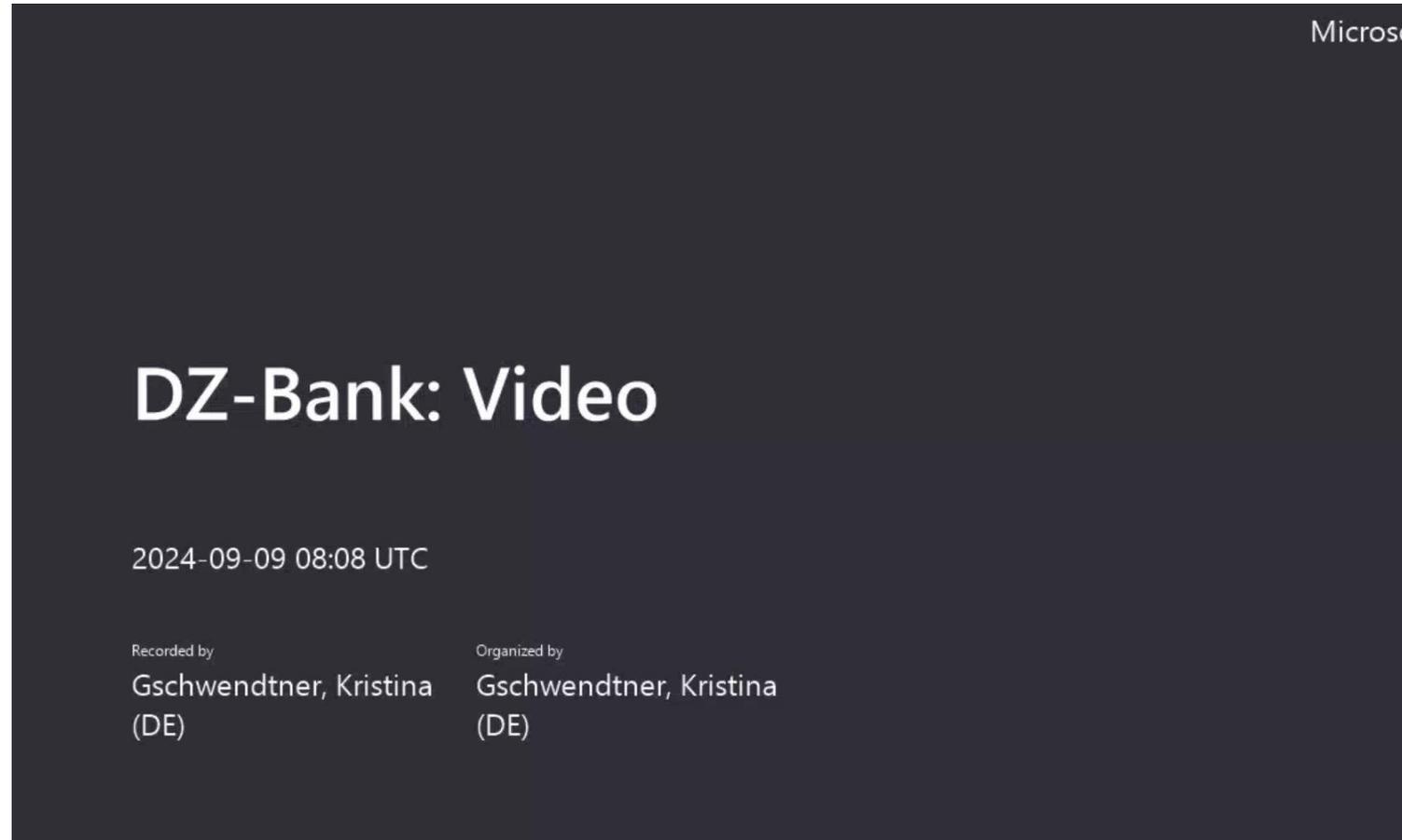
## Qualitätsverbesserung



## Zeitersparnis



# Use Case: Vorbereitung eines Prüfprogramms



Use Case:

Erstellung einer Prozessdokumentation aus  
virtuellen Interviews

# Use Case | Erstellung einer Prozessdokumentation aus virtuellen Interviews

## Kurzbeschreibung

- KI-unterstützte Dokumenten Analyse (Meeting Transkript Zusammenfassung)
- Umwandlung des Texts in Java-Script basierte Mermaid Syntax
- Generierung von Prozess-Visualisierungen in ProtivitiGPT (Vorschau und Download möglich)

## Technologie

- MS Teams
- ProtivitiGPT
- Mermaid Syntax (Java Script basiert)

## Mehrwert

- Enorme Zeitersparnis bei der Erstellung von Meeting-Zusammenfassungen
- Effiziente Erstellung von Prozess-Visualisierungen (Draft-Versionen)

## Komplexität



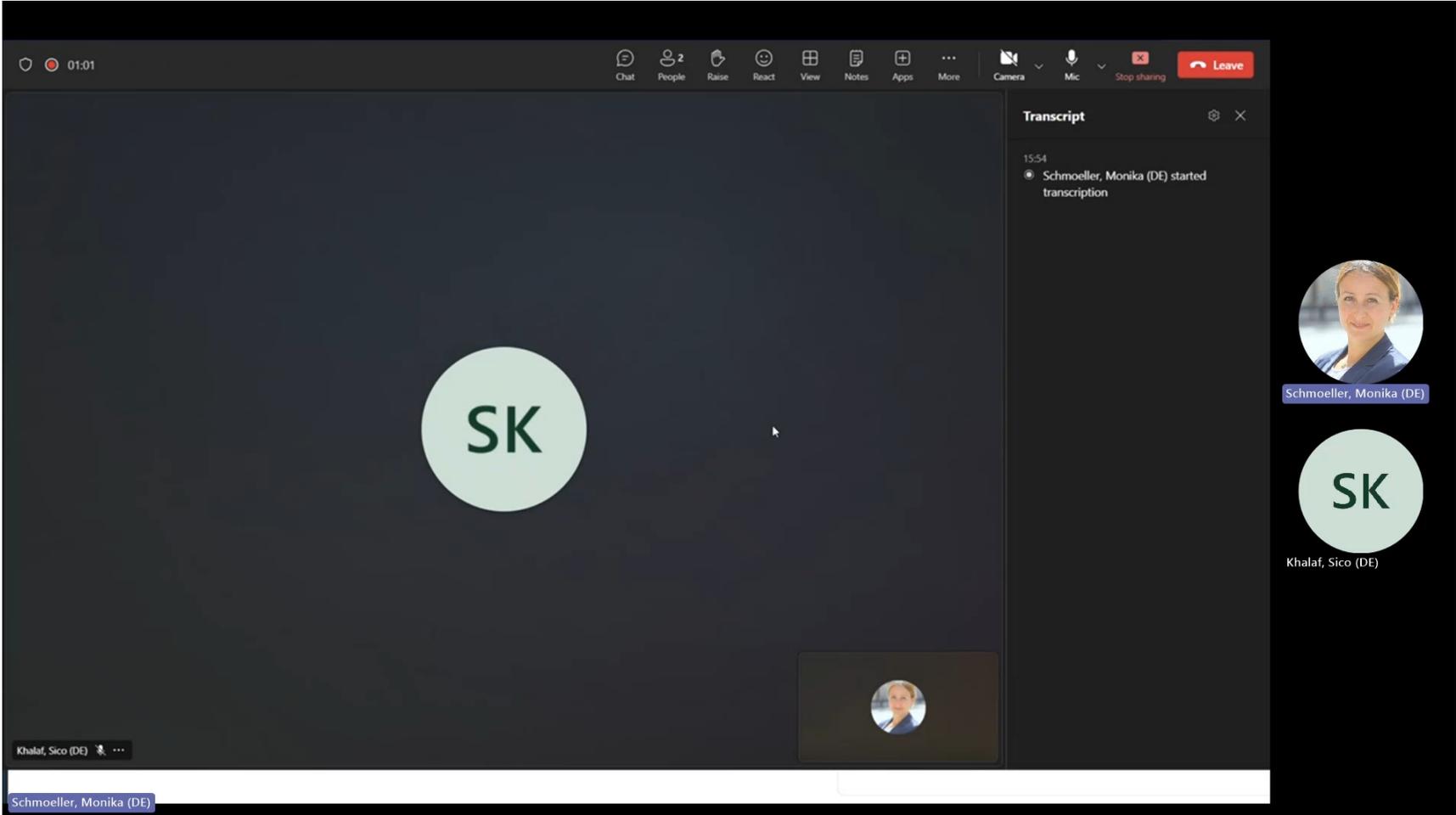
## Qualitätsverbesserung



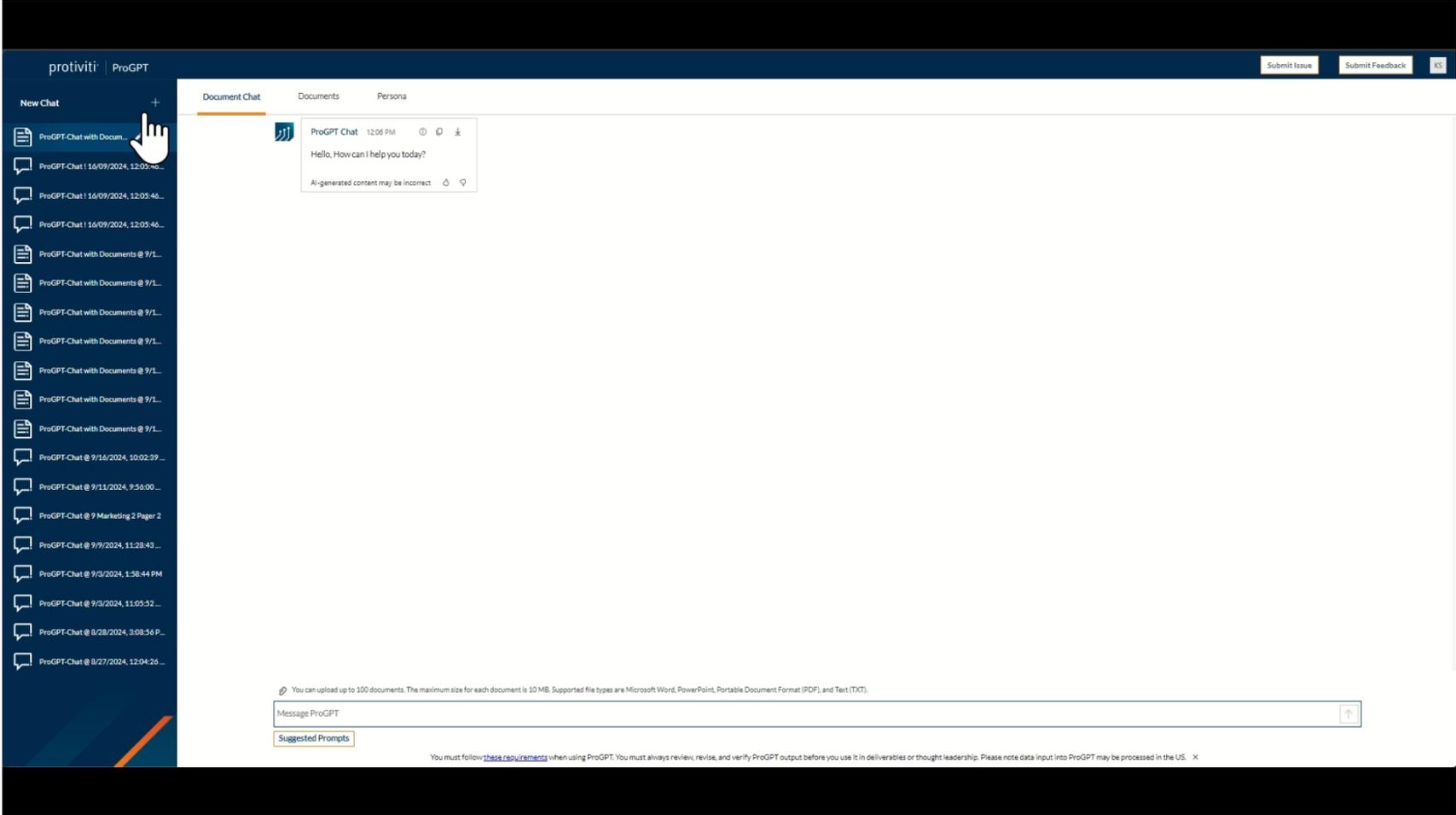
## Zeitersparnis



# ProtivitiGPT | Call Aufzeichnungen



# ProtivitiGPT | Process Flow



Use Case:  
Risikobasierte Sample-Generierung mit GPT

# Use Case: Risikobasierte Sample-Generierung mit GPT

## Kurzbeschreibung

- Hochladen des Datensatzes
- Auswählen des Data Analyst GPT
- Beschreiben der Aufgabe, Ziele und Vorgehensweise
- Schritt für Schritt – Chat zur Ermittlung von Anomalien im Datensatz

## Technologie

- ChatGPT (4o – Data Analyst GPT)  
<https://chatgpt.com/share/81c595e9-559d-4ad9-a9ec-ec209d21bb4a>
- Python, (pandas, sklearn, matplotlib)

## Mehrwert

- Wesentliche Verbesserung des Samples gegenüber einer Zufallsziehung
- Wissensaufbau zu Maschine-Learning-Methoden zur Anomalie-Erkennung

## Komplexität



## Qualitätsverbesserung



## Zeitersparnis



# Use Case: Risikobasierte Sample-Generierung mit GPT – Die Daten

## Credit Card Fraud Detection

Anonymized credit card transactions labeled as fraudulent or genuine



Data Card Code (4896) Discussion (106) Suggestions (0)

### About Dataset

#### Context

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

#### Content

The dataset contains transactions made by credit cards in September 2013 by European cardholders.

This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.

#### Usability ⓘ

8.53

#### License

Database: Open Database, Cont...

#### Expected update frequency

Not specified

#### Tags

Finance

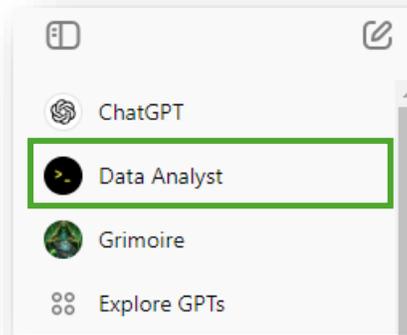
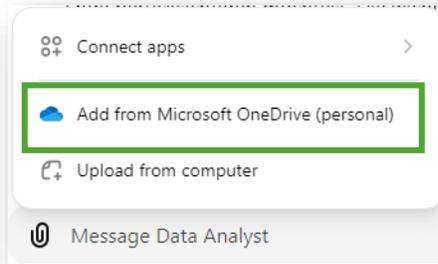
Crime

Quelle:

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

# Use Case: Risikobasierte Sample-Generierung mit GPT

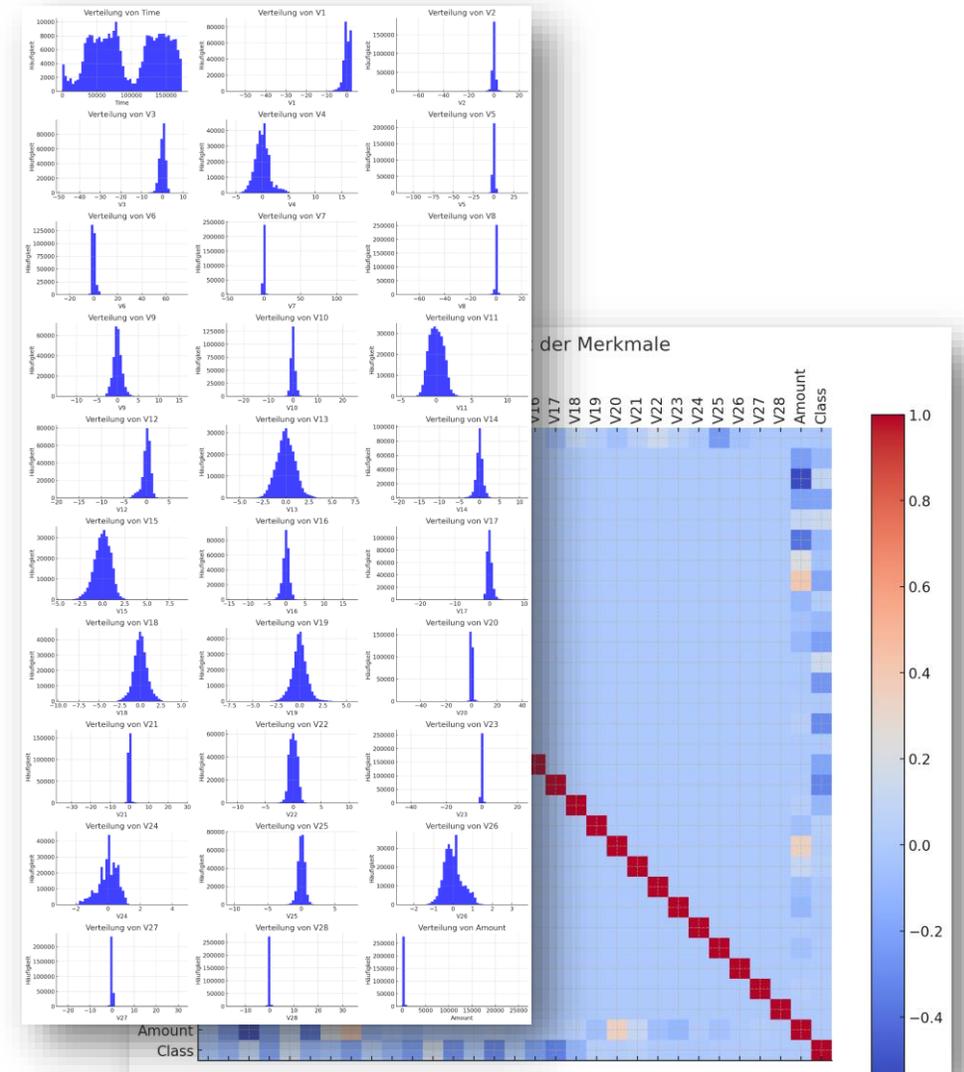
## 1) Rahmen des Chats setzen



Bitte unterstütze mich als Data Scientist. Ich habe einen Datensatz mit Kreditkartenumsätzen und möchte verdächtige Umsätze ermitteln. Der Datensatz enthält eine Spalte „class“ die eine Klassifizierung enthält. Bitte berücksichtige diese Spalte erst am Ende, wenn ich explizit darum bitte. Wenn eine Rechnung wegen Ressourcenmangel abbricht, versuche bitte nicht, das Problem selbst zu korrigieren, sondern nenne jeweils 3 Alternativen mit Vor- und Nachteilen und überlasse mir die Entscheidung, auf welchem Weg wir fortfahren. Hast du Zielsetzung und Ansatz verstanden?

## 2) Verstehen der Daten (Verteilungen, Datenvorbereitung, ...)

Bitte gib mir zunächst eine Übersicht über den Datensatz, bestimme wesentliche Verteilungen der Merkmale, Quantile, Extremwerte, Mittelwerte, Korrelationen und Metriken, die dir sonst noch wichtig erscheinen, den Datensatz zu beschreiben.



# Use Case: Risikobasierte Sample-Generierung mit GPT

## 3) Entscheidung für ein Klassifizierungsverfahren

Ich möchte die Kategorisierung der Daten auf Betrugsverdachtsfälle mithilfe von "unsupervised learning" bestimmen. Welche Klassifizierungsverfahren empfehlst du für die Analyse?



### Zusammenfassung der Empfehlungen:

1. **K-Means oder DBSCAN:** Gut für die grundlegende Clustering-Analyse. DBSCAN eignet sich, wenn du Ausreißer direkt als Anomalien identifizieren möchtest.
2. **Isolation Forest oder LOF:** Besonders geeignet für die Anomalieerkennung und spezifisch darauf ausgerichtet, seltene Ereignisse (wie Betrug) zu identifizieren.
3. **Autoencoder oder GMM:** Eher für komplexere Daten oder wenn du probabilistische Zuordnungen und tiefere Mustererkennung verwenden möchtest.



Es scheint, dass die Berechnung aufgrund der Datenmenge zu lange gedauert hat. Wie besprochen, werde ich dir jetzt drei Alternativen zur weiteren Vorgehensweise vorschlagen:

## 4) Durchführen der Klassifizierung

Kannst du bitte den Datensatz mit Hilfe des Isolation Forest auf Anomalien untersuchen.

## 5) Anpassung und Fehlerbehebung

Kannst Du bitte das Model mit der Hälfte der Bäume berechnen, um Rechenzeit zu sparen.

- 1.425 verdächtige (anomale) Transaktionen

Wenn du möchtest, können wir die identifizierten Anomalien weiter analysieren oder visualisieren.

Lass es mich wissen, wie du fortfahren möchtest! [\[>-\]](#)

```
# Isolation Forest-Modell mit reduzierter Anzahl an Bäumen (n_estimators)
iso_forest_reduced = IsolationForest(contamination=0.01, n_estimators=50, random_state=42)

# Modell auf die Daten anpassen
iso_forest_reduced.fit(data_features)

# Vorhersage der Anomalien (1 = normal, -1 = Anomalie)
anomalies_reduced = iso_forest_reduced.predict(data_features)

# Anomalien in den Datensatz integrieren
data['Anomaly'] = anomalies_reduced

# Anzahl der Anomalien und normalen Punkte
anomaly_counts_reduced = data['Anomaly'].value_counts()

anomaly_counts_reduced
```

# Use Case: Risikobasierte Sample-Generierung mit GPT

## 6) Analyse der Anomalien

Welche Merkmale sind wichtig für die Anomalie-Erkennung?

Wie visualisiere ich Anomalien?

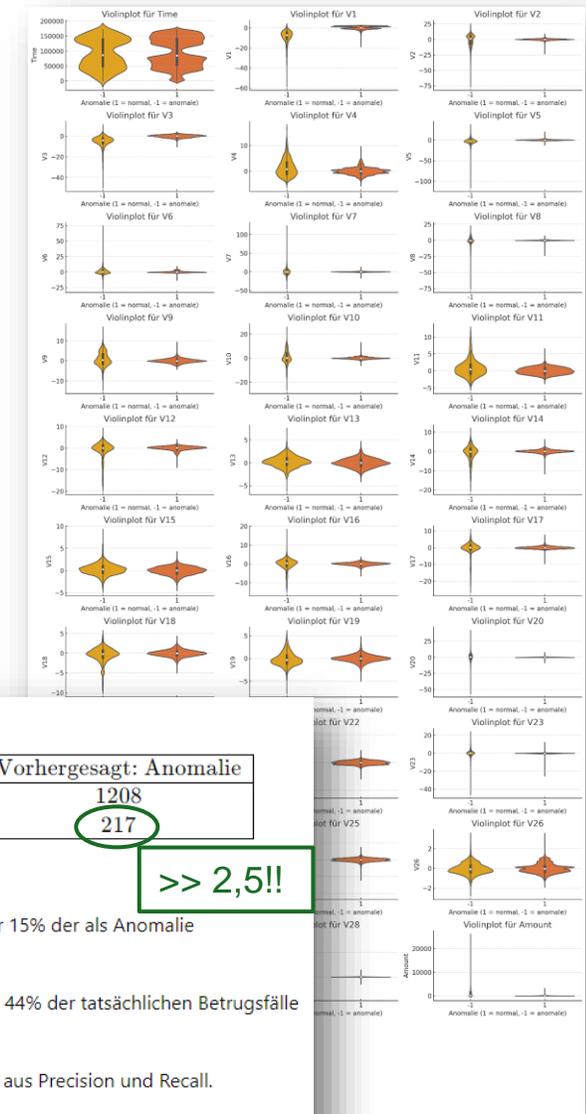
Visualisiere die Anomalien in einem Violinplot pro Merkmal. (anstelle der vorgeschlagenen Boxplots)

### Mögliche Ansätze zur Identifizierung wichtiger Merkmale:

1. Shapley-Werte (SHAP)
2. Univariate Analyse
3. Feature Contribution Analysis

### Hier sind einige gängige Ansätze zur Visualisierung von Anomalien:

1. PCA-Reduktion auf 2D oder 3D
2. t-SNE (t-Distributed Stochastic Neighbor Embedding)
3. Scatterplot von 2 Features mit Anomalie-Kennzeichnung
4. Boxplots der Merkmale für normale und anomale Daten



## 7) Bewerten der Ergebnisse

Lass uns dieses Ergebnis als Endergebnis definieren! Bitte analysiere die Qualität/Performance des Modells mittels einer Konfusionsmatrix. Ermittle bitte zusätzliche Kennzahlen, um die Qualität des Modells zu bewerten. Ab jetzt kannst du natürlich die Spalte "class" verwenden.

### Konfusionsmatrix:

	Vorhergesagt: Normal	Vorhergesagt: Anomalie
Tatsächlich: Normal	283107	1208
Tatsächlich: Anomalie	275	217

### Bewertung der Modelleistung:

- **Precision (Genauigkeit der Anomalie-Erkennung): 15%** – Nur 15% der als Anomalie identifizierten Transaktionen sind tatsächlich Betrug.
- **Recall (Erkennungsrate der tatsächlichen Anomalien): 44%** – 44% der tatsächlichen Betrugsfälle wurden korrekt erkannt.
- **F1-Score: 23%** – Der F1-Score ist ein harmonischer Mittelwert aus Precision und Recall.
- **Genauigkeit (Accuracy): 99%** – Das Modell ist insgesamt sehr genau, da die meisten Transaktionen normal sind, aber die Performance bei der Anomalie-Erkennung ist eher gering.

>> 2,5!!

## 8) Speichern des Samples

Stell die Liste der Anomalien als Download zur Verfügung.

Die Liste der als Anomalien identifizierten Transaktionen v erstellt. Du kannst die Datei über den folgenden Link herunterladen:

[Anomalien herunterladen](#)

Lass es mich wissen, wenn du weitere Analysen oder Anpassungen wünschst! [-]

Use Case:  
Entwicklung eines Berichtsentwurfs aus  
Prüfungsartefakten

# Use Case: AI Audit Reporting

## Kurzbeschreibung

- Excellisten mit den Prüfungsartefakten (z.B. Feststellungen, Prüfungsprogramm, Prüfungsscope) hochladen
- Ausgabeparameter festlegen (z.B. Observation Refinement)
- Bericht generieren

## Technologie

- ProtivitiGPT
- MS PowerApp
- MS Excel

## Mehrwert

- Große Zeitersparnis bei der Erstellung eines Prüfungsberichts
- Effizienz des Revisors steigern

## Komplexität



## Qualitätsverbesserung



## Zeitersparnis



# Use Case: AI Audit Reporting

## protiviti® | AI Use Case Reporting Tool

### Introduction

The AI Use Case Reporting tool will use a large language model (LLM) to produce a good quality audit report based on the information within certain inputted audit files.

The tool aims to help Protiviti employees and clients not only improve the quality of the audit reports they create, but also drastically reduce the time it takes to write them.

The outputted audit report will include two main sections - the Executive Summary and the Detailed Findings & Recommendations.

### What you will need

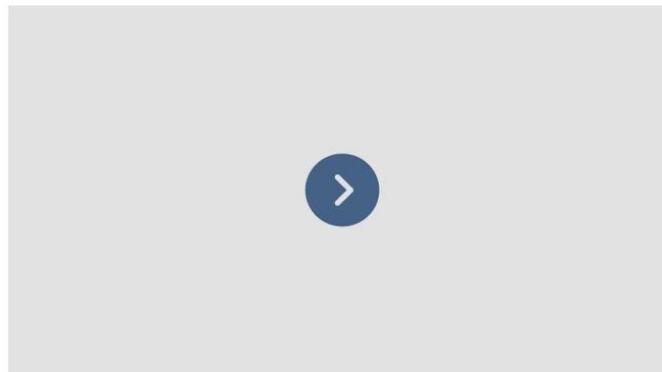
Please be ready to submit the following audit files for your project:

- Scoping Memo
- Work Program / RCM
- Observations Log

The files and data you provide will be contained within Protiviti's secure IT environment. Unless you are restricted by the terms of the client contract or by Protiviti-provided guidance, you can input client-specific information.

### What is TAP?

The Protiviti Technology Accelerator Platform (TAP) is a self service application that provides users from all solutions the ability to reduce time and effort on tedious tasks via automation.



### Templates

These templates are available for use at your convenience. If you already have a version of the documents that aligns with the requirements, you are welcome to use those instead.

Observations Log Template [Download](#)

Work Program Template [Download](#)

Scope Memo Template [Download](#)

### Detailed Instructions

1. Check that each the input files cover the following pieces of information:
  - Scoping Memo
    - Background on the audit project
    - Objectives & scope
    - Key milestone activities of the project
    - Time period under review
  - Work Program / RCM
    - Conclusions on the effectiveness of each control
  - Observations Log
    - Every observation noted throughout the audit project
    - The risk rating for each observation
2. Submit each of the three required files
3. Indicate whether you'd like any of the following items within the audit report to be further refined by the tool, or kept the same as shown in the observations log. This refinement would involve integration of the 5 C's of audit reporting and other best report writing practices. The items are:
  1. Observations language
  2. Risk statement of each observation (this would be generated by the LLM if not already indicated in the observations log)
  3. Recommendation for each observation (this would be generated by the LLM if not already indicated in the observations log)

[Begin](#)

# Risiken und Regulatorik

# Definition von KI-Modellen

## EU AI ACT

---

„KI-System“ bezeichnet ein maschinelles System, das für den Betrieb mit unterschiedlichen Autonomiegraden ausgelegt ist und nach der Bereitstellung möglicherweise anpassungsfähig ist und das aus den eingehenden Daten ableitet, wie es aus diesen Daten Ergebnisse wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen generieren kann, die physische oder virtuelle Umgebungen beeinflussen können, und zwar für explizite oder implizite Ziele.

---

## MaRisk

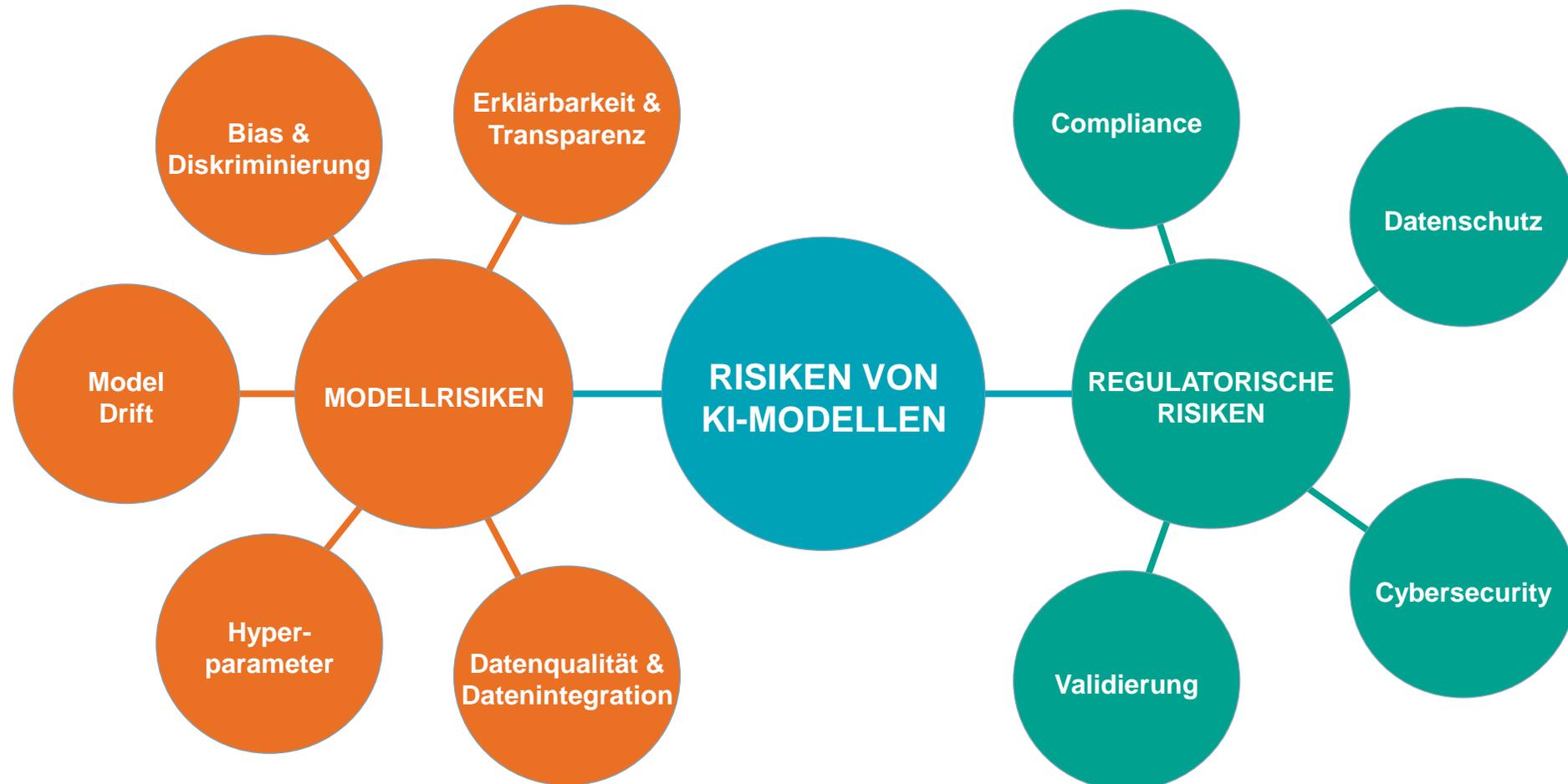
---

Ein Modell im Sinne dieses Moduls ist eine quantitative Methode, ein System oder ein Ansatz, der statistische oder mathematische Theorien, Techniken und Annahmen anwendet, um Eingabedaten zu quantitativen Schätzungen zu verarbeiten.

---

# Risiken von KI-Modellen

Ethische KI basiert auf Transparenz, Verantwortlichkeit und Vertrauen und bedeutet zu verstehen, wie KI-Algorithmen entwickelt werden, welche Daten zum Trainieren der Maschinen verwendet werden und welche Auswirkungen dies auf Unternehmen, Menschen und Gemeinschaften hat.



# Best Practices

# Best Practices der Anwendung von Data Analytics/ KI im Revisionsumfeld

## Strategien & Richtlinien

### Strategien

- Zielsetzung definieren
- Stakeholder einbinden
- Marktanalysen durchführen
- **Pilotprojekte definieren**
- Skalierbarkeit beachten

### Richtlinien

- Compliance
- Datenschutz und Sicherheit
- Ethik und Transparenz
- Qualitätskontrolle

## Prozesse

- **Zielgerichtete Anwendungsfälle**
- Integration in bestehende Systeme
- Workflow-Optimierung
- Echtzeit-Datenverarbeitung
- Kundenerlebnis verbessern
- Feedback-Mechanismen
- Regelmäßige Evaluierung
- Change-Management
- Dokumentation und Richtlinien

## Menschen & Organisation

- **Kultur des Wandels fördern**
- **Schulung und Weiterbildung**
- Kommunikation und Transparenz
- Interdisziplinäre Teams bilden
- **Rollen und Verantwortlichkeiten klar definieren**
- Engagement der Führungsebene sicherstellen

## Methoden

- **Inkrementelle Einführung**
- Agile Methodik anwenden
- **Partnerschaften mit Technologieanbietern**
- Intensive Datenaufbereitung
- Verwendung erprobter Algorithmen
- Regelmäßige Evaluierung und Anpassung der Modelle

## Systeme & Daten

- Robuste IT-Infrastruktur
- **Datensicherheit gewährleisten**
- Datenintegration optimieren
- Skalierbarkeit berücksichtigen
- Data Governance Rahmenwerk etablieren
- Qualitätsmanagement für Daten

# Q&A

*Face the Future with Confidence*<sup>®</sup>

© 2024 Protiviti – Confidential. Protiviti is not licensed or registered as a public accounting firm and does not issue opinions on financial statements or offer attestation services. All registered trademarks are the property of their respective owners.

**protiviti**<sup>®</sup>  
Global Business Consulting