

FROM DATA TO DECISIONS: UNLEASHING THE POWER OF GEN AI

Versicherungsmathematisches Kolloquium der LMU München

15.07.2024

Amelie Morgenstern, Tom Huber, Daniel Hohmann

ENTSTEHUNG DES HEUTIGEN VORTRAGTITELS

Dear AI-chatbot,
Give me a funny or catchy title for a presentation on use cases of generative AI in the insurance business

Of course! Here are a few ideas for a funny or catchy title for your presentation on use cases of generative AI in the insurance business:

- "From AI to LOL: How Generative AI is Revolutionizing Insurance"
- "Insuring the Future: The Hilarious Side of Generative AI"
- "AI-nsurance: Laughing All the Way to the Bank with Generative AI"
- "The AI Insurance Comedy Show: Unleashing the Power of Generative AI"
- "Generative AI: The Insurance Industry's Secret Weapon for Laughs and Success"

Feel free to mix and match or modify these suggestions to suit your presentation style and audience. Good luck with your presentation!

INHALT

01

Einführung in die Welt der KI

02

Large Language Models

03

Generative Adversarial Networks

01

EINFÜHRUNG IN DIE WELT DER KI

DIE WELT DER KÜNSTLICHEN INTELLIGENZ



Künstliche Intelligenz (KI) / Artificial Intelligence (AI) ist ein Zweig der Informatik, der sich mit der Schaffung und Entwicklung von Maschinen befasst, die wie Menschen denken und handeln.

Maschinelles Lernen (ML) ist ein Programm oder System, das anhand von Eingabedaten ein Modell trainiert, das dann anhand neuer oder noch nie zuvor gesehener Daten nützliche Vorhersagen treffen kann. ML gibt dem Computer auch die Fähigkeit zu lernen, ohne explizit zu programmieren.

Neuronale Netze (NN) sind vom menschlichen Gehirn inspiriert und ahmen die Art und Weise nach, wie sich biologische Neuronen Signale verarbeiten und senden. NN bestehen aus miteinander verbundenen künstlichen Neuronen, die in Schichten organisiert sind.

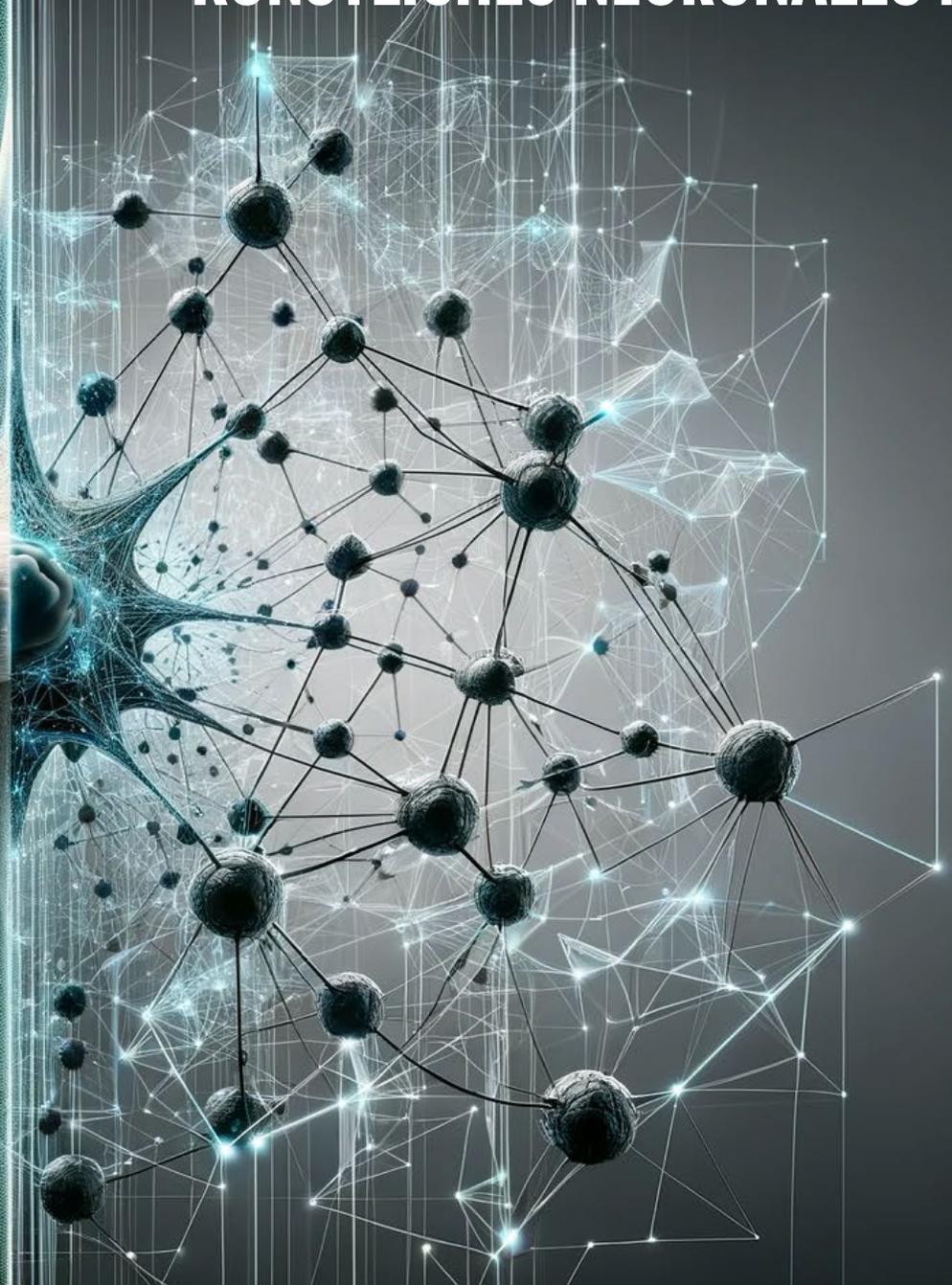
Deep Learning (DL) bezieht sich auf jedes neuronale Netz mit mehr als drei verborgenen Schichten. Mit einer höheren Anzahl versteckter Schichten sind DL-Modelle gut geeignet, um komplexe reale Probleme anzugehen.

Generative KI (GenAI) ist eine Untergruppe des DL, die auf der Grundlage dessen, was aus vorhandenen Inhalten gelernt wurde, verschiedene Arten von Inhalten - wie Texte, Bilder, Audio, Videos - generieren kann.

NEURONALES NETZ DES GEHIRNS



KÜNSTLICHES NEURONALES NETZ



NEURONALE NETZE

Was ist das eigentlich?

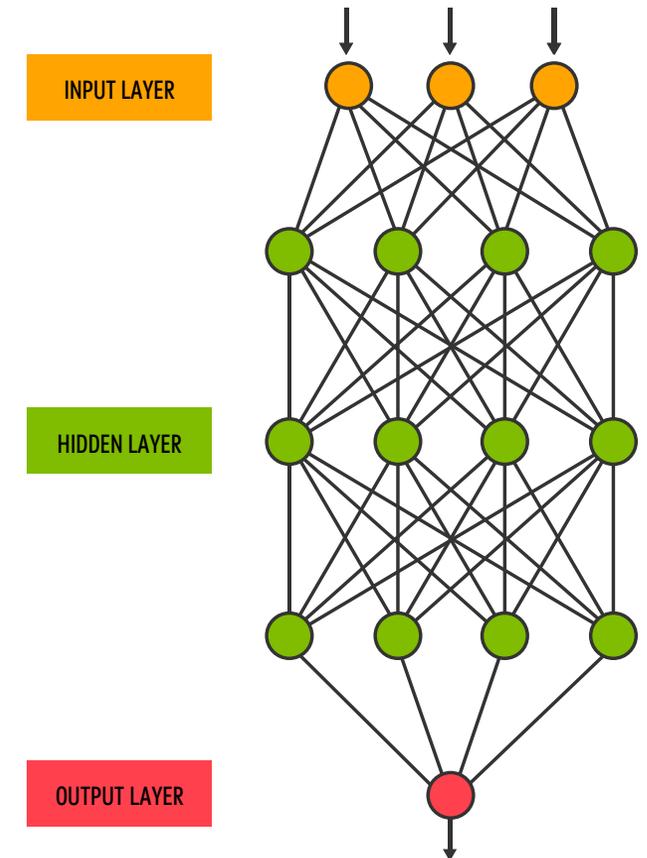
THEORETISCHER HINTERGRUND

- **Inspiziert von biologischen Systemen:** Ein neuronales Netzwerk ist eine Methode der künstlichen Intelligenz, die Computern beibringt, Daten auf eine Art und Weise zu verarbeiten, die dem menschlichen Gehirn nachempfunden ist. Es schafft ein lernfähiges System, das es Computern ermöglicht, aus ihren Fehlern zu lernen und sich ständig zu verbessern. Ziel beim Einsatz von Neuronale Netzen ist es, komplizierte Probleme wie die Zusammenfassung von Dokumenten oder die Erkennung von Gesichtern mit größerer Genauigkeit zu lösen.
- **Neuronen und Schichten:** Ein neuronales Netz besteht aus unterschiedlichen Schichten (layer) von Knoten oder "Neuronen". Jedes Neuron empfängt die Signale, die von der vorherigen Schicht gesendet wurden, verarbeitet diese und gibt ein neues Signal an die nächste Schicht weiter. Die erste Schicht ist die Eingabeschicht, die letzte Schicht ist die Ausgabeschicht, und jede Schicht dazwischen wird als versteckte Schicht (hidden layer) bezeichnet.
- **Komplexe Abhängigkeiten:** Neuronale Netze können Computern helfen, intelligente Entscheidungen mit begrenzter menschlicher Unterstützung zu treffen, da sie sehr komplexe, nichtlineare Beziehungen zwischen Eingabe- und Ausgabedaten erlernen und modellieren können.
- **Lernprozess:** Neuronale Netze lernen, indem sie ihre Gewichte (Parameter) als Reaktion auf Trainingsdaten oder andere Arten von Feedback anpassen. Dies geschieht in der Regel durch einen Prozess namens Backpropagation, bei dem der Gradient einer Verlustfunktion in Bezug auf die Gewichte berechnet wird, so dass diese schrittweise in Richtung eines geringeren Verlusts angepasst werden können.
- **Deep Learning:** Neuronale Netze mit vielen verborgenen Schichten werden als tiefe neuronale Netze bezeichnet, und die Erforschung dieser Netze nennt man Deep Learning. Dieses hat sich in vielen Bereichen bewährt, z.B. bei der Bild- und Spracherkennung, der Verarbeitung natürlicher Sprache, usw.



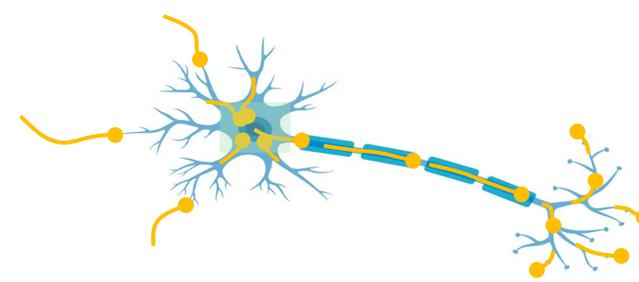
Komplexität und Tiefe des NN haben erhebliche Auswirkungen auf den Umfang der abgedeckten Hypothesen (möglichen Ausprägungen des NN)

DARSTELLUNG EINES NEURONALEN NETZES



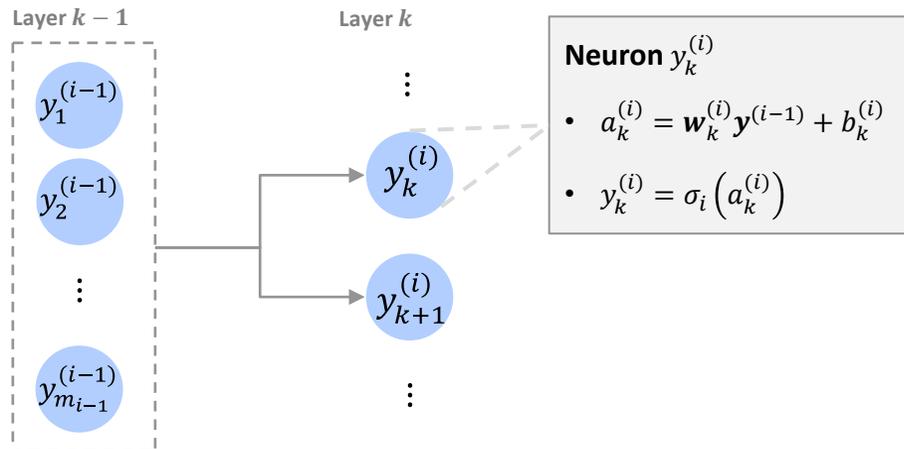
NEURONALE NETZE

Neuronen und Aktivierungsfunktion



NEURONEN

- $y_k^{(i)}$: Ausgabe des Neurons k der Schicht i , $y_k^{(0)} := x_k$
- m_i : Anzahl Neuronen in der Schicht i
- $\mathbf{w}_k^{(i)}, b_k^{(i)}$: Parameter des Neurons k der Schicht i
- σ_i : Aktivierungsfunktion der Schicht i



Die Berechnungen innerhalb jedes Neurons sind sehr einfach und machen das gesamte NN mathematisch verhältnismäßig leicht zugänglich

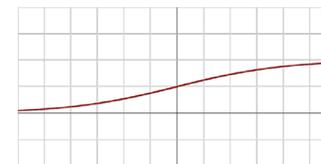
AKTIVIERUNGSFUNKTION

- Aktivierungsfunktionen dienen dem Zweck, Nichtlinearitäten abzubilden:

$$\mathbf{y}^i = \sigma_i(\mathbf{W}^i \sigma_{i-1}(\mathbf{W}^{i-1} \mathbf{y}^{i-2} + \mathbf{b}^{i-1}) + \mathbf{b}^i)$$

- Aktivierungsfunktionen können sich von Schicht zu Schicht unterscheiden
- Die Aktivierungsfunktion der Ausgangsschicht hat erhebliche Auswirkungen auf das Verhalten des Ausgangssignals

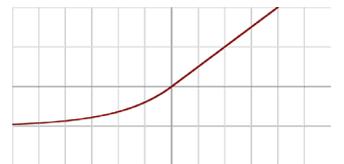
Sigmoid



ReLU



ELU



Aktivierungsfunktionen sind das Schlüsselement für die Beschreibung komplexer Interaktionen und Abhängigkeitsstrukturen

MODELLE FÜR MASCHINELLES LERNEN

Im Allgemeinen nimmt die

Prognosekraft des maschinellen Lernmodells ab

Hoch

Interpretierbarkeit

Niedrig

Methode/Klassen

Bayesianisch

Logistische Regressionen

Regressionen

Clustering

Entscheidungsbaum

Support-Vektor-Maschinen

Neuronale Netze

Ensemble-Lernen

Neuronale Netze

Mehrschichtige neuronale Netze (Deep Learning)

Ensemble-Methoden

DAV

DEUTSCHE AKTUARVEREINIGUNG e.V.

Ergebnisbericht der Ausschüsse Rechnungslegung und Regulierung und Actuarial Data Science

Regulierung und Validierung von KI-Modellen

Köln, 26. Februar 2024

Netz aus logistischen Regressionen (oder ähnlichen Klassifikatoren) mit einer versteckten Schicht

Netzwerk von Klassifikatoren mit mehreren ausgeblendete Schichten

Kombination von mehreren Modellen in Reihe

Wann wir es verwenden können

Klassifikator unter Annahme der

ator

Wenige Klassifikatoren und Regression (LASSO, Ridge, Elastic Net)

Regression führt eine Variablenauswahl überanpassung

Kernzentren nach dem Zufallsprinzip die Punkte in diesen Clustern gruppiert werden. Wenn eine Konvergenz erreicht ist

Wählt ein Cluster. Die Anzahl der Cluster wird die nächstgelegenen Cluster

Entscheidungsbaum-Klassifikatoren, die nach falsch gewichtet werden

Entscheidungsbaum-Klassifikatoren, die auf zufälligen Merkmalen trainiert wurden

Support-Vektor-Maschine, aber in der Regel

Wenige Merkmale und nicht linear trennbar

Wenige Merkmale und hochkomplexe, nichtlineare Trennung

Falls die Fehleranalyse von anderen Modellen auf zusätzliche Leistungsgewinne hindeuten

Erstellen einer Basisprognose

Klassifizierung linear trennbarer Daten

Modellierung linear trennbarer Daten (oder nichtlineare Wechselwirkungen können leicht aufgebaut werden)

Nichtlineare Trennung mit vielen Merkmalen

Gewöhnliche Clustering-Probleme (Kunden, Geschäfte, Regionen, Produkte, ...)

Scheitert bei nicht-linearen Datensätzen

Clustering-Problem: Kunden, Geschäfte, Regionen, Produkte.. Scheitert bei nicht-linearen Datensätzen

Nichtlineare Trennung mit viele Merkmale

Nichtlineare Trennung mit vielen Merkmalen

Nichtlineare Trennung mit wenigen Merkmalen und vielen Datenpunkten

Rechnerischer Aufwand

Sehr niedrig

Niedrig

Niedrig

Niedrig

Niedrig

Mittel

Mittel

Mittel

Hoch

Hoch

Sehr hoch

Hoch - Sehr hoch

WAS HAT SICH GEÄNDERT?

3 FAKTOREN:

BEISPIEL:



GRÖÖÖÖ



Zunahme der Rechenleistung
zur Beschleunigung der
Datenverarbeitung

GPUs: erfordert 30.000 Nvidia
Grafikkarten zur Verarbeitung

TRAINING



Verbesserte Algorithmen zur
Verarbeitung immer
komplexerer Datenmengen

Input: trainiert auf 300 Mrd.
Wörtern

GELD

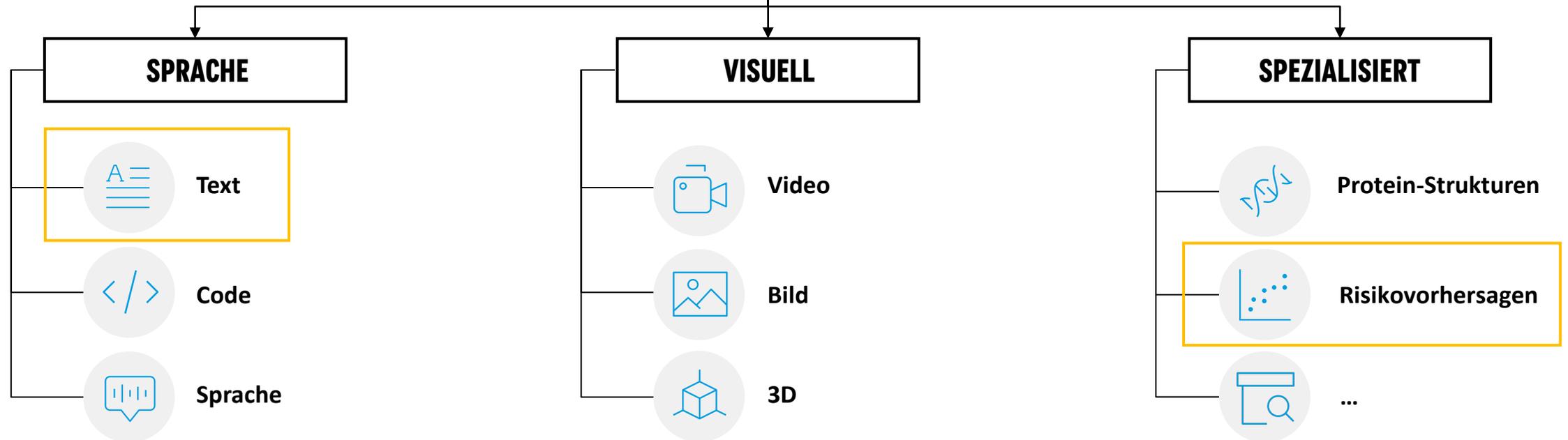


Aufschwung der Finanzierung
ermöglicht Investitionen in Zeit
und Geld

Ungefähre **Kosten** für einen
Trainingslauf: 100 Millionen
Dollar

GEN AI MODELLE KÖNNEN WEIT MEHR ALS NUR TEXT ERZEUGEN ...

GEN AI MODALITÄTEN



02

LARGE LANGUAGE MODELS



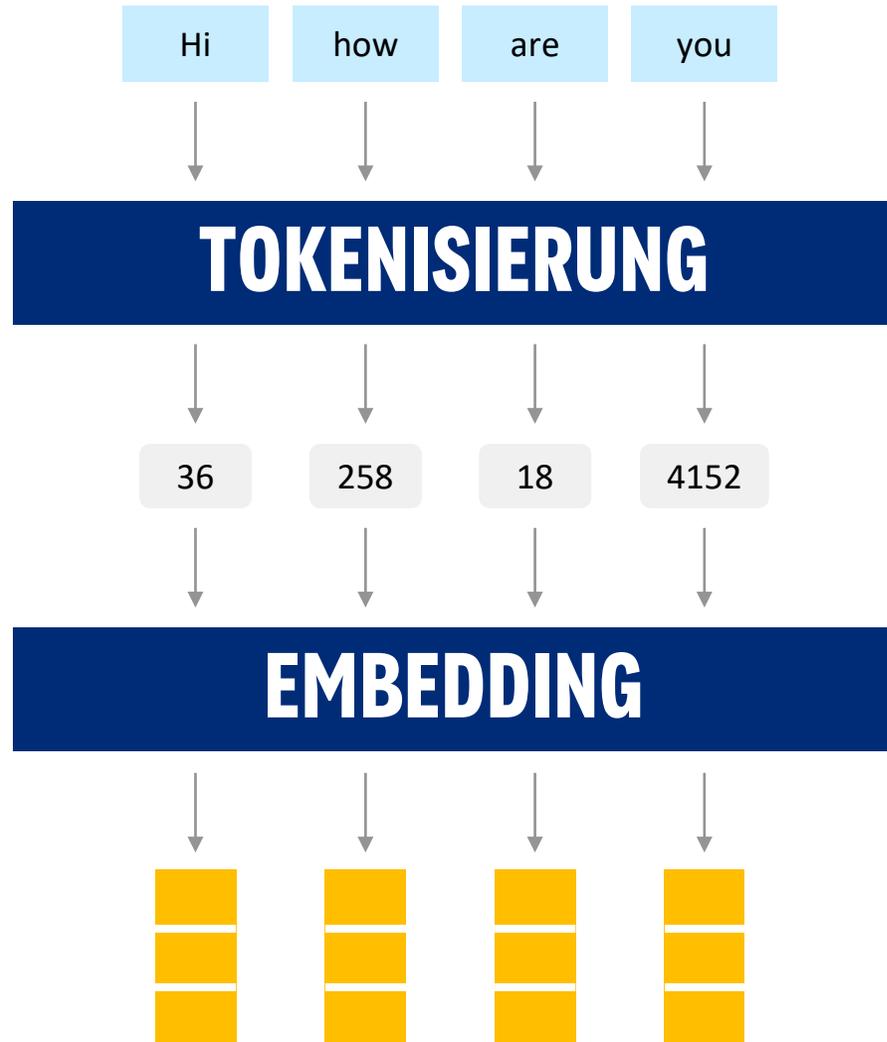
**FORTSCHRITTE DURCH
EINE ARCHITEKTUR
NAMENS
„TRANSFORMER“
VORANGETRIEBEN,
MIT FOLGENDEN
KOMPONENTEN...**

A Embedding

B Positional encoding

C Self-attention

EMBEDDING HILFT BEI DER ÜBERSETZUNG VON MENSCHLICHER SPRACHE IN MASCHINENSPRACHE



Übersetzung von Wörtern in eindeutige numerische Werte:

1 Wort \geq 1 Token

Umwandlung von Token in hochdimensionale Vektoren (bis zu 512 Dimensionen)

$$H_i = \begin{matrix} 0.10 \\ 0.54 \\ 0.29 \end{matrix}$$

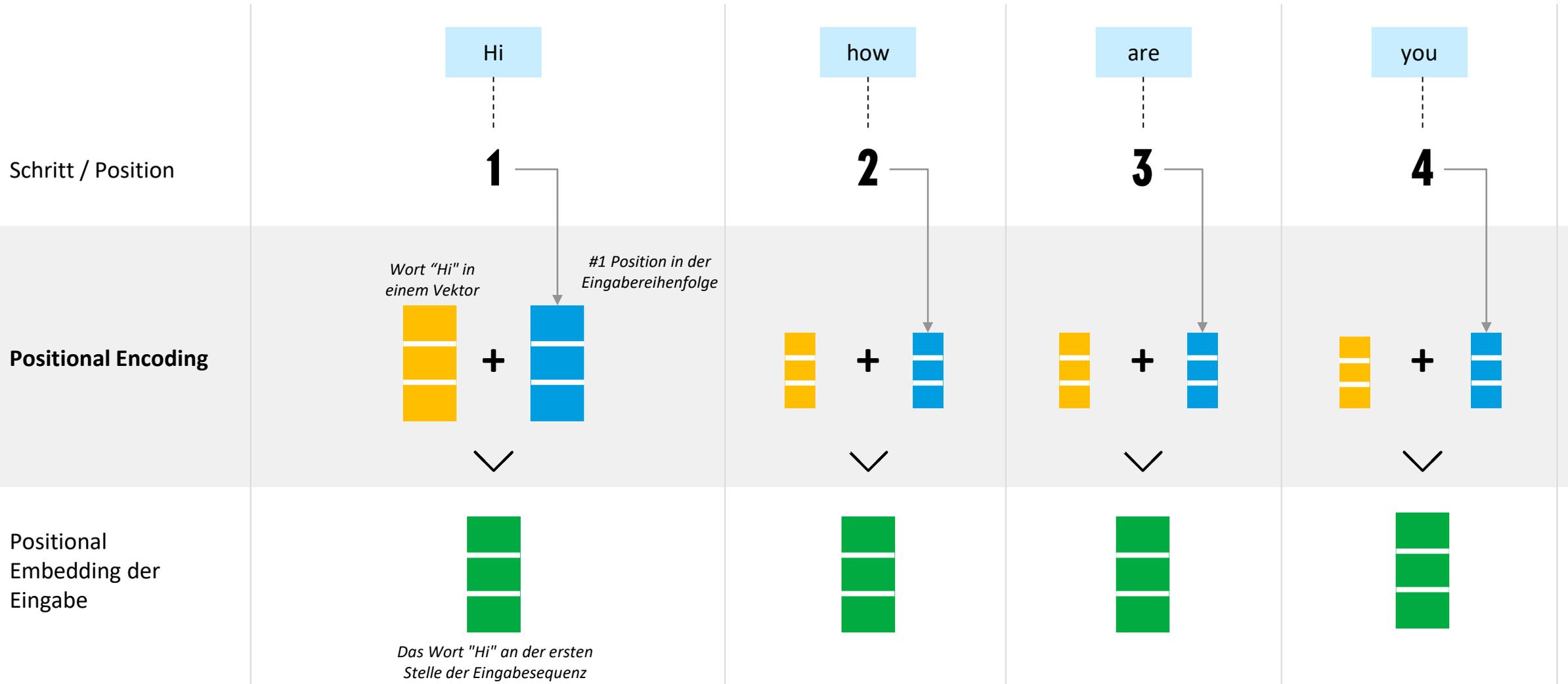
Vektor

DIE VEKTORISIERUNG SPEICHERT ÄHNLICHE WÖRTER NAHE ZUSAMMEN



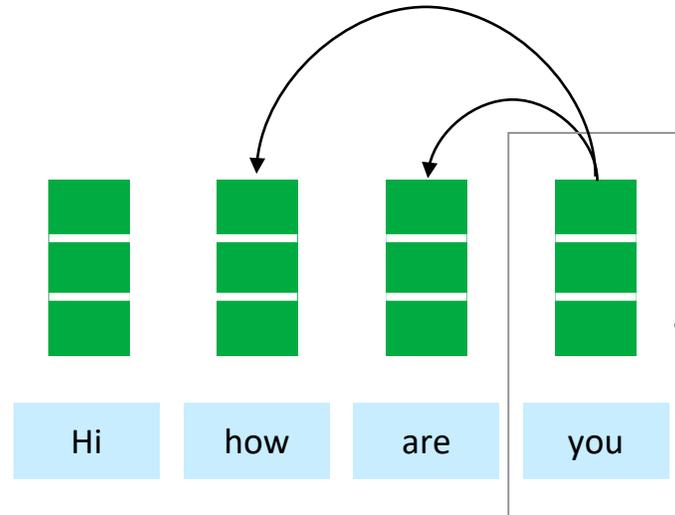
> **Kitten, Dogs** und **Puppy** kommen oft im gleichen Kontext wie das Wort **Cat** vor

DAS POSITIONAL ENCODING UMFASST DIE WORTPOSITION IM VEKTOR



SELF-ATTENTION ERSCHLIEßT DAS INNOVATIVE POTENZIAL VON GENAI ZUR KONTEXTUALISIERUNG VON WÖRTERN

Assoziation des Wortes "You" mit anderen Wörtern des Satzes



MECHANISMUS DER SELF-ATTENTION:

Den Kontext des Wortes innerhalb des Satzes verstehen

	Hi	how	are
you	0.1	0.4	0.2

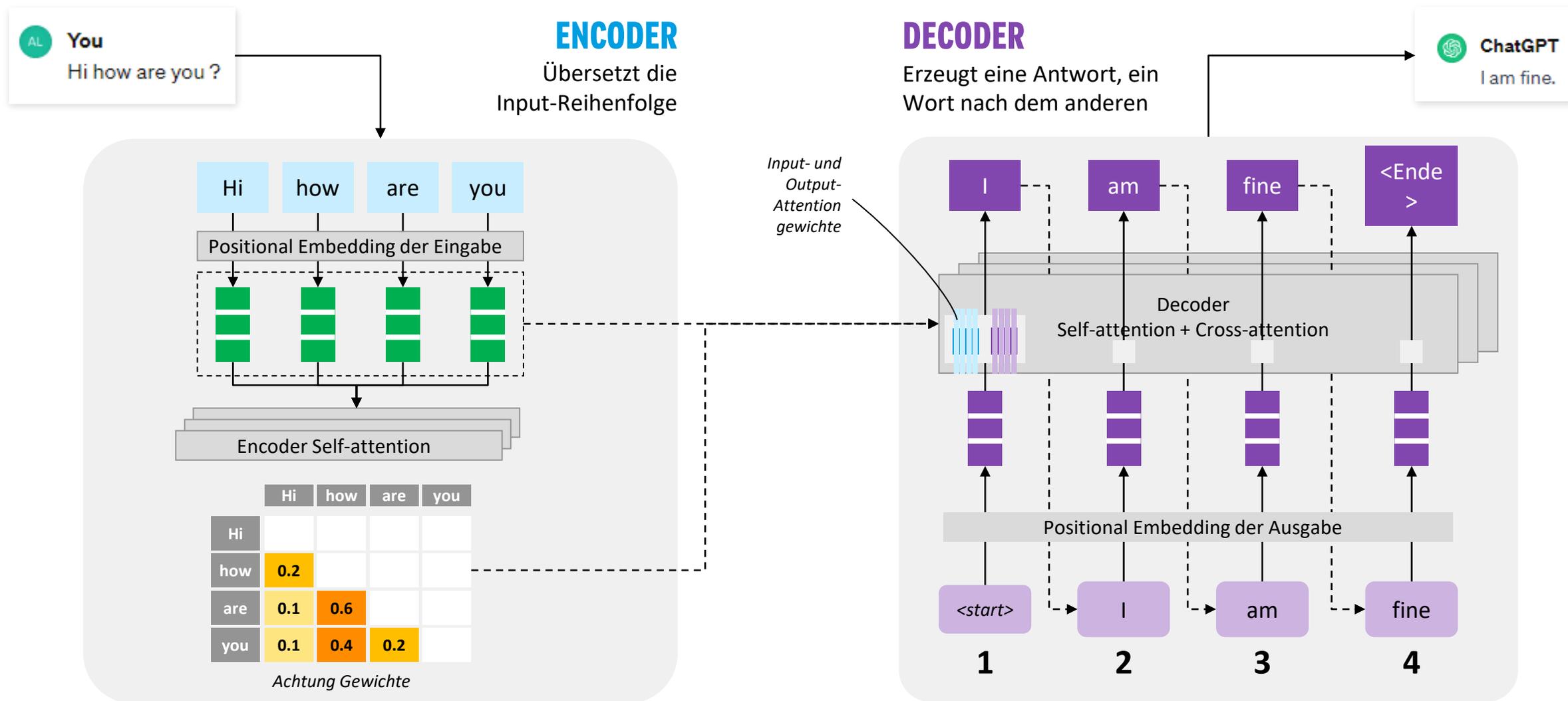
Achtung Gewichte

Matrix, die die Bedeutung der Beziehungen zwischen den Wörtern in der Eingabesequenz bestimmt



Einfacher und schneller zu trainieren als ältere Modelle (rekurrente neuronale Netze)

SEHEN WIR UNS AN, WIE DIE TEXTERSTELLUNG FUNKTIONIERT



WAS SIND DIE RISIKEN UND GRENZEN VON LLMS?

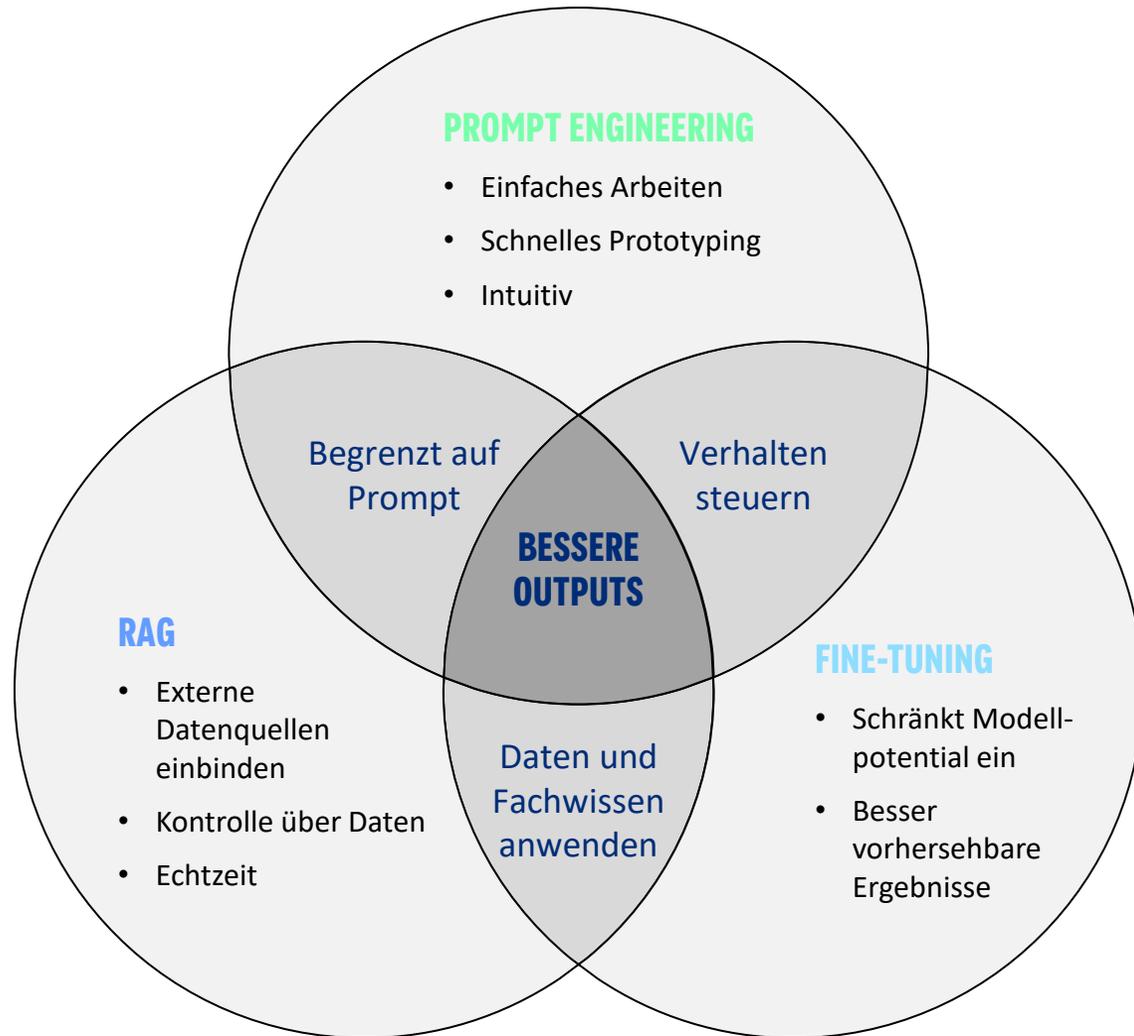


Technologische Risiken

Risiken der Verwendung

 *Schwerpunkt der nächsten Folie*

OPTIMIERUNGSMÖGLICHKEITEN FÜR LLMS:



OPTIMIERUNGSMÖGLICHKEITEN:

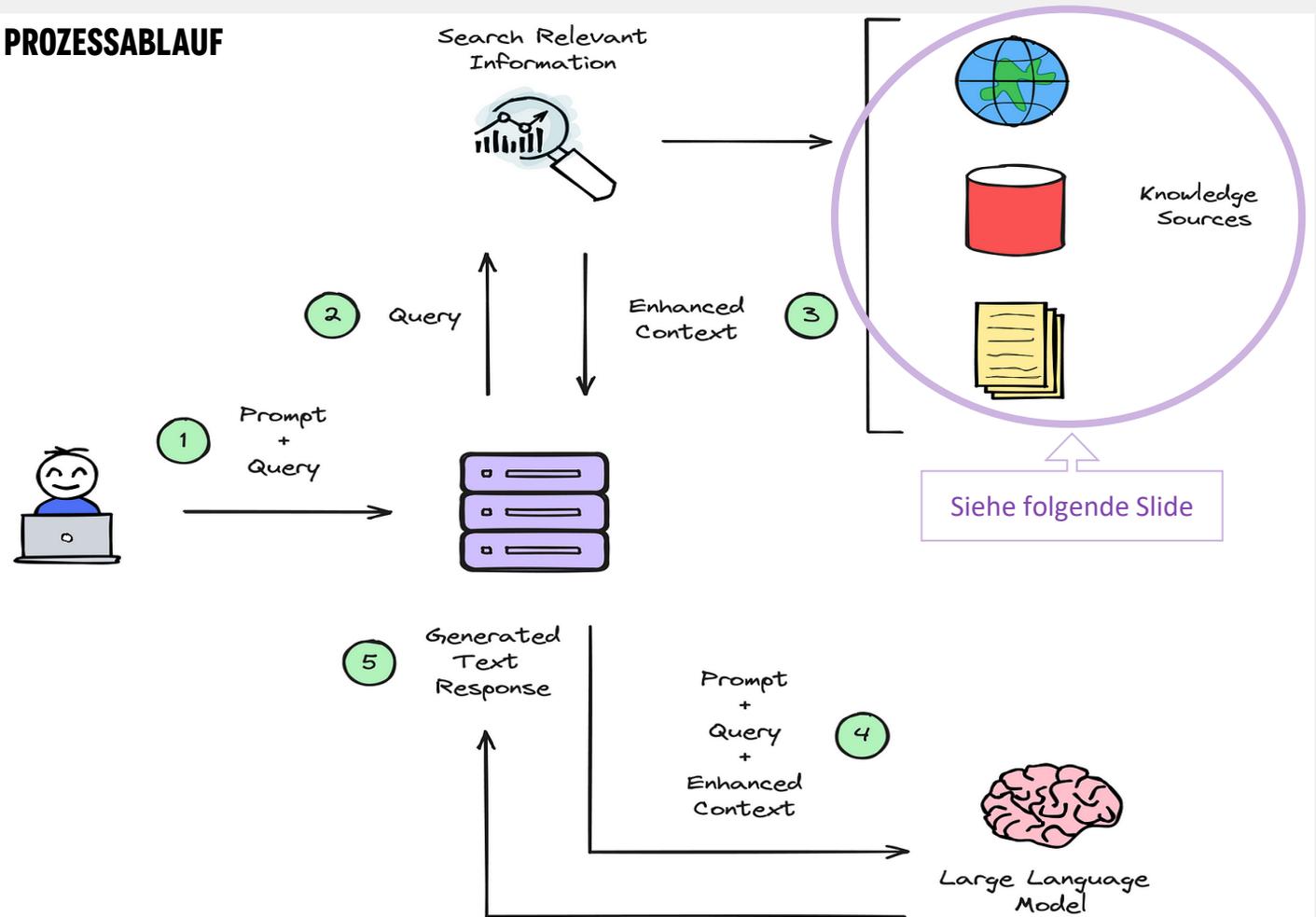
- **PROMPT ENGINEERING** ist der Prozess, bei dem durch die Auswahl geeigneter Formate, Ausdrücke, Wörter und Symbole die generative künstliche Intelligenz (generative KI) gesteuert wird, um die gewünschten Ergebnisse zu erzielen und eine sinnvolle Interaktion mit Benutzern zu ermöglichen.
- **RETRIEVAL-AUGMENTED GENERATION (RAG)** ist der Prozess der Optimierung der Ausgabe eines LLMs, so dass es zur Generierung einer Antwort auf eine externe Wissensbasis außerhalb seiner Trainingsdatenquellen zugreift.
- **FINE-TUNING** ist ein Prozess, bei dem ein vortrainiertes LLM auf einen spezifischen Anwendungsfall trainiert wird. Dies ermöglicht eine bessere Leistung und Anpassungsfähigkeit des Modells für den spezifischen Anwendungsfall.

IDEE VON RETRIEVAL AUGMENTED GENERATION

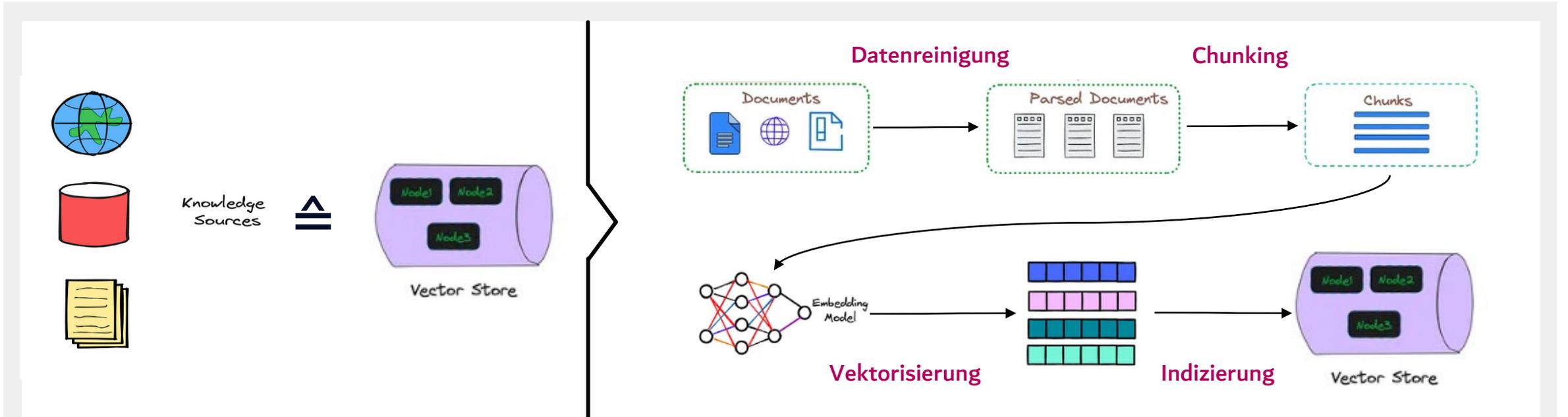
RETRIEVAL AUGMENTED GENERATION (RAG)

- Retrieval Augmented Generation (RAG) ist ein innovativer Ansatz, der Elemente der **traditionellen Spracherzeugung** mit Techniken des **Information Retrieval** kombiniert.
- RAG integriert eine Retrieval-Komponente in den Generierungsprozess, die es dem Modell ermöglicht, während der Textgenerierung auf **externe Wissensquellen zuzugreifen**.
- Die Retrieval-Komponente ruft relevante Informationen aus **großen Wissensdatenbanken auf der Grundlage des dem Modell bereitgestellten Kontexts** ab.
- Diese abgerufenen Informationen dienen als **zusätzlicher Input oder Kontext für das Generierungsmodell** und **verbessern die Relevanz, Kohärenz und Genauigkeit** des generierten Textes.

RAG - PROZESSABLAUF



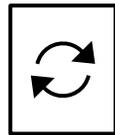
RAG VECTOR STORE



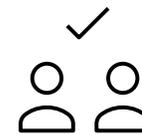
RAG VORTEILE



KOSTENGÜNSTIGE UMSETZUNG



AKTUELLE INFORMATIONEN



ERHÖHTES VERTRAUEN DER NUTZER
DURCH QUELLENANGABEN



MEHR KONTROLLE FÜR ENTWICKLER

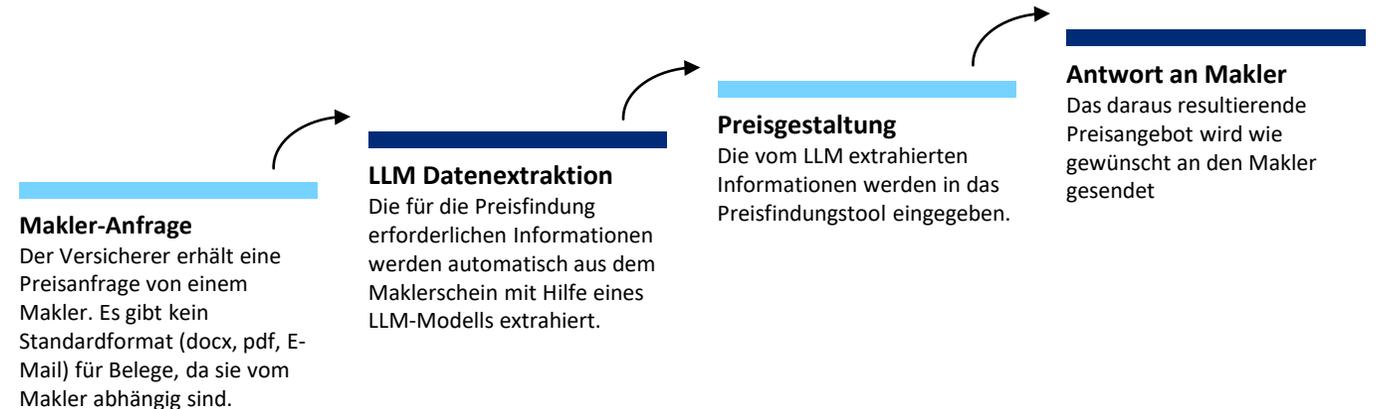
FALLSTUDIE 1: LLMS UND DATA MINING

Effizienzgewinne durch automatisierte Erstellung von Underwriting-Angeboten

Hintergrund und Einblicke

- **Hintergrund:** Die Extraktion von Informationen aus unstrukturierten Daten, z. B. manuelle Erstellung von Underwriting-Angeboten aus Makleranfragen ist zeitaufwändig und fehleranfällig.
- **Idee:** Automatisierung des Prozesses mit Hilfe von LLMs:
 - Extrahieren Sie Schlüsselinformationen wie z.B. Details zur Deckung durch ein LLM.
 - Standardisieren und formatieren Sie die Informationen zu einheitlichen Makleranfragen.
 - Anpassung des Angebots an Kundenwünsche und Branchenstandards.
 - Erleichterung der Zusammenarbeit, der Überprüfung und der Genehmigungsverfahren.
 - Integration in bestehende Systeme zur Automatisierung und Effizienzsteigerung.
- **Potenzielle Gefahr:** Fehler bei der Datenextraktion oder -interpretation

Pipeline für die automatische Erstellung von Underwriting-Angeboten



```
prompt = f"""What is the identification code and the lob mentioned in the below? Which company is it?
What's the email address and the phone number of the sender?

{source_document}

Provide the output using the following structure:
[
  identification code;,
  lob;,
  company;,
  email address;,
  phone number;,
]"""

print(llm.invoke(prompt).content)
```



```
identification_code: UK8727180938Z
lob: Flexible Life Plan
company: IBrokerage
email_address: Paul.Smith@ibrokerage.com
phone_number: +44 6549 000000

** Process exited - Return Code: 0 **
Press Enter to exit terminal
```

FALLSTUDIE 2: REGULATORISCHE UND INTERNE MODELLSUCHE

Hintergrund und Einblicke

- **Hintergrund:** Modelle können auf relevante Themen abgestimmt werden oder mit einer vollständigen, eingebetteten Bibliothek von Dokumenten versehen werden, z. B. regulatorische oder interne Modelldokumentationen
- Ein Nutzer kann nach Inhalten suchen, die Antwort enthält:
 1. Zusammengefasste Antwort(en)
 2. Rückverweise auf Quelldokumente
- Bewertung der Antwortqualität
- Zu den Erweiterungen, die dieselben Modelle verwenden, gehören
 - Identifizierung von Überschneidungen oder Widersprüchen in den verschiedenen regulatorischen Dokumenten
 - Ergänzung oder Ersatz von Guidelines
 - Überwachung und Zusammenfassung von Aktualisierungen der Rechtsvorschriften in verschiedenen Rechtsordnungen

Screenshots aus dem OW-internen Tool zur Abfrage und Suche von Dokumenten

Welche Regelungen sind bei der Anwendung einer Matching-Anpassung zu berücksichtigen?

Welche Annahmen liegen dem derzeitigen Ansatz zur Modellierung des Zinsrisikos in unserem internen Modell zugrunde? Bitte stellen Sie die Ergebnisse in einer übersichtlichen Tabelle zusammen.

94% Our confidence rating is 94%

Based on:

- ✓ We were able to understand your question very well
- ✓ Sufficient documents were found relating to your query, which you have permission to view
- ✓ The answer provided was rated high in relevance and coherence to your question
- ✓ We found little ambiguity and uncertainty around the possible answers to your question

2. Data
2.1. Data Collection: We collect...

OliverWyman IntelliQA

Ask a Question

Can we keep a customer email address indefinitely?

We retain personal data for as long as individuals remain subscribed to our email marketing list or until they choose to unsubscribe. When an individual unsubscribes, their personal data is promptly deleted from our email marketing system.

OliverWyman

ESMA

ESMA-2014-180 EN

Guidelines on the use of internal models

Opinion on the supervisory aspects including a dynamic volatility

Legal basis

1. The European Insurance and Occupational Pensions Authority (EIOPA) issues this opinion on the basis of Article 29(1)(a) of Regulation (EU) No 1094/2010. According to this article EIOPA shall play an advisory role and shall not exercise supervisory powers. This opinion is issued in accordance with the supervisory culture and consistent supervisory practices of the Union.
2. This opinion is based on Directive 2009/138/EC (Solvency II Directive) and Commission Delegated Regulation (EU) 2015/35 (Solvency II Delegated Regulation) and other relevant instruments.
3. This opinion is addressed to the competent authorities of the Union in accordance with Article 4(2) of Regulation (EU) No 1094/2010.

Context and objective

4. EIOPA is attentive to the convergence of models and underlying compliance with the Solvency II Directive, further developed and supplemented in EIOPA's guidelines.
5. The volatility adjustment (VA) is one of the key elements of the Solvency II value-at-risk (VaR) framework. It aims at stabilising the Solvency II VaR in periods of high market volatility by adding an extra risk charge to the calculation of technical provisions.
6. Concerning internal models, there are two approaches that take the VA into account (SCR) by allowing the VA to move in line with the 1-year forecast of basic own funds. In the approaches the VA is kept constant in the SCR calculation.

ESMA - Weithofen Tower | Westhafenplatz 1 - 60327 Frankfurt - Germany - Tel. + 49 69 959-10000
Fax: + 49 69 959-10001 | e-mail: info@esma.europa.eu | web: https://esma.europa.eu

PROJECT GROUP

Study on Diversification in Internal Models

Public Report

24 January 2024

ESMA

FALLSTUDIE 3: AUTOMATISCHE ERSTELLUNG VON BERICHTSENTWÜRFEN

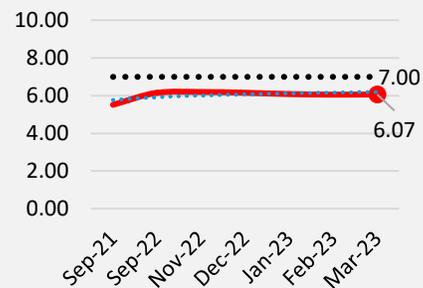
Effizienzsteigerungen durch die Erstellung von Entwürfen für Diagramme und Texte in allen Berichten, aufbauend auf Automatisierung, Datenkonsolidierung und Implementierung von mehrstufigen Dashboards

Hintergrund und Einblicke

- Einsatz generativer KI zur Erstellung von Entwürfen von Schlussfolgerungen und zur Erstellung eines ersten Drafts für Berichte für Führungskräfte
- LLMs automatisieren die Auswertung großer Mengen statischer und Zeitreihen Daten, um Erkenntnisse und Diagramme zu erstellen
- Die Ausgaben umfassen
 - Ausführliche Zusammenfassungen
 - Regelmäßige Statusberichte
 - Identifizierung von Themen/Trends
 - Verweise auf Datenquellen
 - Qualitätsbewertungen für das Vertrauen in die erzielten Ergebnisse

Metrikname und Werte als Eingabe für das LLM

% der automatisierten Kontrollen



% der präventiven Kontrollen



Automatisch generierte Erkenntnisse mit LLM API

Daten zu Erkenntnissen für einzelne Metriken

① In diesem Zeitraum ist der Prozentsatz der automatisierten Schlüsselkontrollen stetig, aber leicht angestiegen, und zwar von 5,52 % im September 2021 auf 6,07 % im März 2023. Trotz dieses Anstiegs bleibt der Prozentsatz durchweg hinter der Zielvorgabe zurück, was darauf hindeutet, dass das Unternehmen seine Automatisierungsanstrengungen möglicherweise beschleunigen muss, um das festgelegte Ziel von 7 % automatisierter Schlüsselkontrollen zu erreichen.

② Während dieses Zeitraums zeigt der Prozentsatz der präventiven Schlüsselkontrollen einen allmählichen Rückgang von 57,99 % im September 2021 auf 54,93 % im März 2023. Außer im September 2021 blieb der Prozentsatz der präventiven Schlüsselkontrollen unter der Zielvorgabe, was auf ein potenzielles Gap bei der Erreichung des gesetzten Ziels hindeutet. Dieser Abwärtstrend und die konsequente Unterschreitung der Zielvorgabe lassen darauf schließen, dass die Organisation ihre präventiven Kontrollmaßnahmen bewerten und verbessern muss.

Automatisch generierte Zusammenfassung unter Verwendung von Spracheingabe und LLM API

Individuelle Einblicke in Exec Summary

The control environment in the company demonstrates significant improvements in control quality, as evidenced by a downward trend in the percentage of controls rated unsatisfactory from September 2021 to February 2023. Moreover, the effectiveness of key controls consistently remains above the minimum limit, reflecting strong control practices within the organization.

①

+

②

+

•

•

•

⊗

In terms of automation, a steady but slow increase is observed in automated key controls, although these percentages consistently fall short of the target. Semi-automated controls, despite showing a downward trend, remain consistently above the threshold, revealing a nuanced scenario in the semi-automation of key controls.

03

GENERATIVE ADVERSARIAL NETWORKS

EINFÜHRUNG IN GENERATIVE ADVERSARIAL NETWORKS (GAN)

Hintergrund

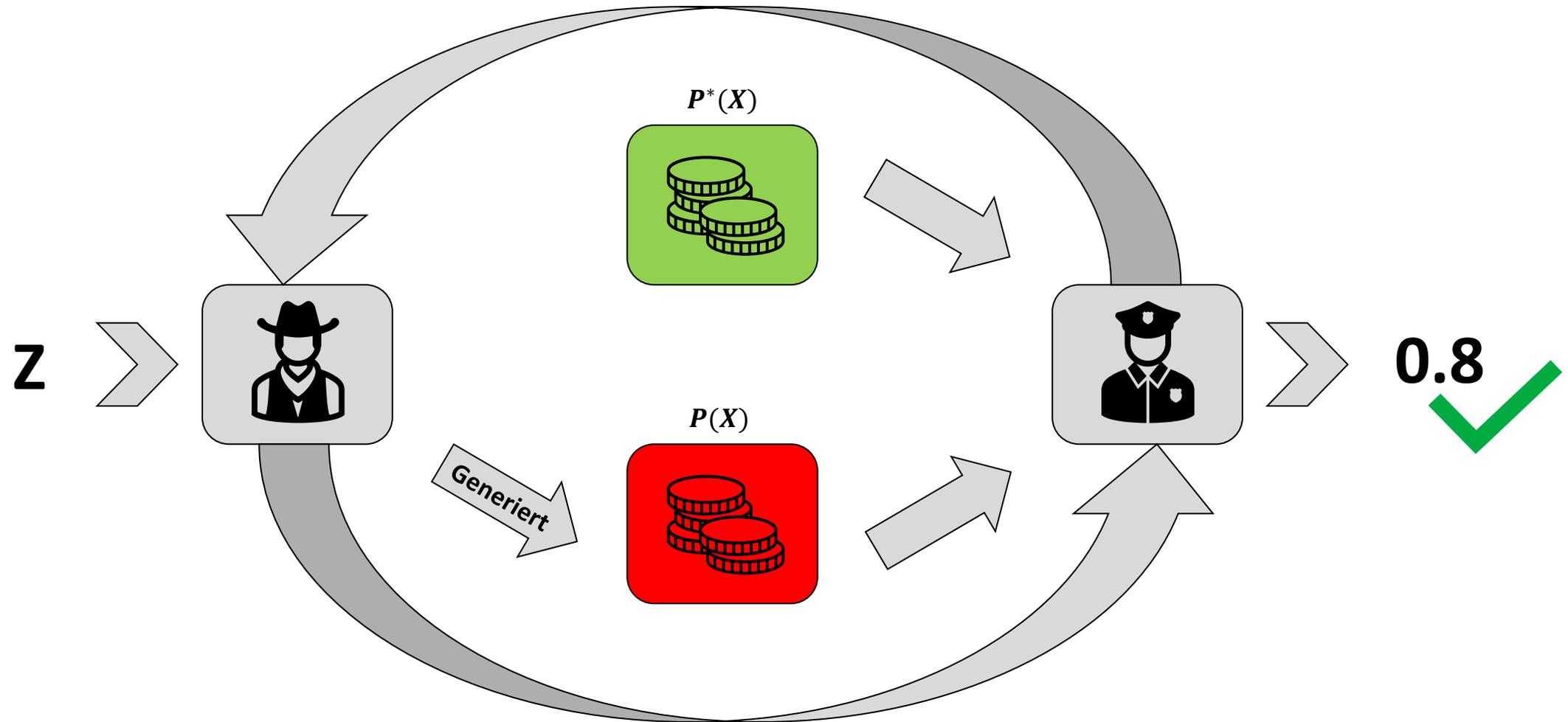
- Die generative Modellierung ist ein Bereich des maschinellen Lernens, der sich mit der Modellierung einer über Datenpunkte \mathbf{X} definierten Verteilung $P(\mathbf{X})$ in einem potenziell hochdimensionalen und komplexen Raum befasst
- Wir können diesen Aufbau formalisieren, indem wir sagen, dass wir Beispiele \mathbf{X} erhalten, die entsprechend einer unbekanntem Verteilung $P^*(\mathbf{X})$ verteilt sind, und unser Ziel ist es, ein Modell für \mathbf{P} zu lernen, von dem wir eine neue Stichprobe \mathbf{X}^* ziehen können, deren Verteilung so ähnlich wie möglich zu P^* ist.
- Eine in jüngerer Zeit entwickelte Technik zur generativen Modellierung **ist das so genannte** generative adversarial network (GAN), das im Zusammenhang mit der Bilderzeugung entwickelt wurde und hier auf die Marktrisikomodellierung angewandt wird, bei der es darum geht, anhand einer Verteilung zu simulieren, die die historische Entwicklung der Marktrisikofaktoren so gut wie möglich darstellt, und eine beliebige Anzahl von Stichproben zu ziehen (die den historischen Risikofaktorwerten ähnlich, aber nicht mit ihnen identisch sind).



Herausforderungen:

- Der Trainingsalgorithmus bringt eine komplexe Dynamik mit sich, da sich intuitiv sowohl der Diskriminator als auch der Generator "Hand in Hand" bewegen müssen, um pathologische Lösungen zu vermeiden

GANS: SPIELTHEORETISCHE INTERPRETATION



$$\min_{\theta_G \in \Theta_G} \max_{\theta_D \in \Theta_D} E[\log(D_{\theta_D}(X))] + E[\log(1 - D_{\theta_D}(G_{\theta_G}(Z)))]$$

FALLSTUDIE 4: MODELLIERUNG VON REAL-WORLD RISIKOFKTFOR VERTEILUNG

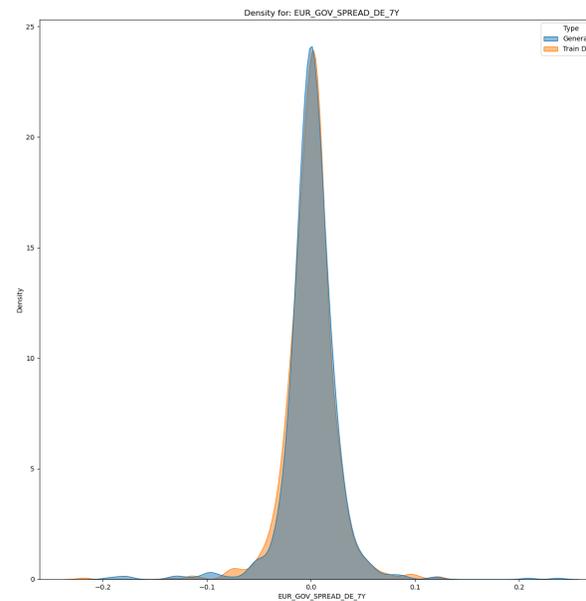
Idee & Umsetzung

Motivation:

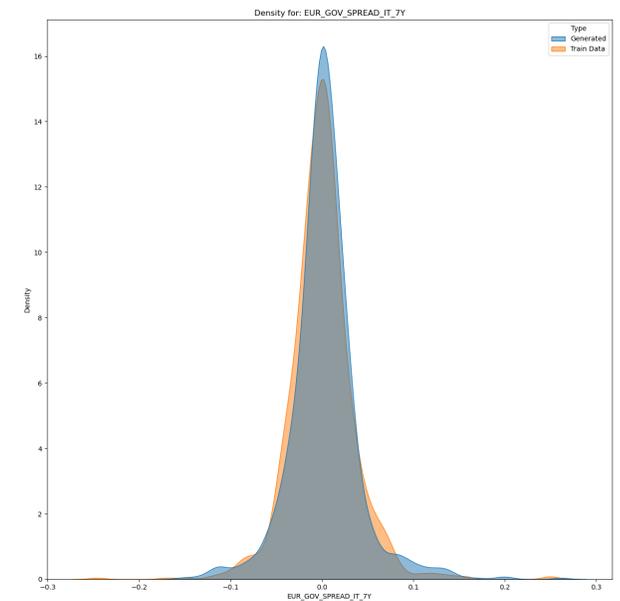
- Die Modellierung von Risikofaktoren in der realen Welt ist komplex und umfasst in der Regel die Kalibrierung von
 - geeigneten marginalen (univariaten) Risikofaktorverteilungen und
 - einer Kopula, die die gesamte Abhängigkeitsstruktur widerspiegelt.
- Die Vielzahl parametrischer Annahmen bezüglich der gemeinsamen Wahrscheinlichkeitsverteilung führt häufig zu einer qualitativ unbefriedigenden Wiedergabe des historisch beobachteten Verhaltens ökonomischer Variablen.
- Strukturelle Abhängigkeiten (z. B. Safe-Haven-Effekte) können in der Regel nicht durch stochastische Modelle erfasst werden

Beispiel:

- Lernen einer 10-dimensionalen multivariaten Verteilung (wöchentliche log-returns)
- Darunter EQ, EQ-Vol, Spread und Swap

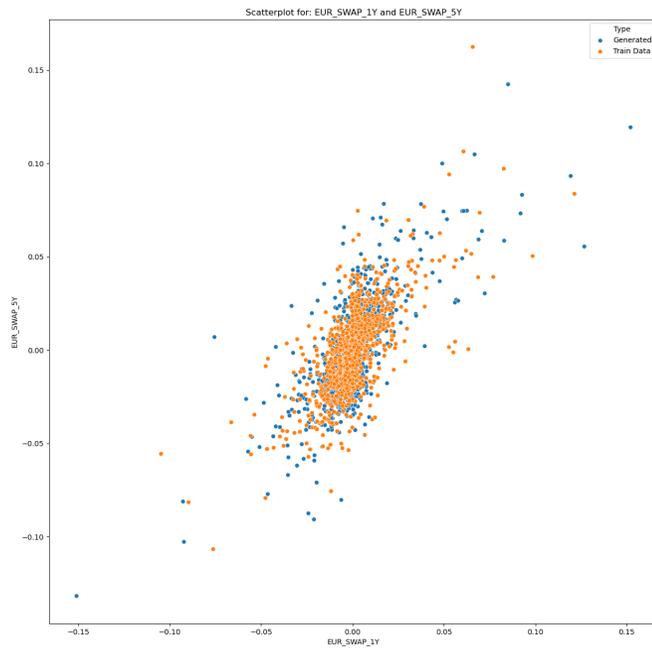


Marginalverteilung: DE Gov-Spread (7Y)

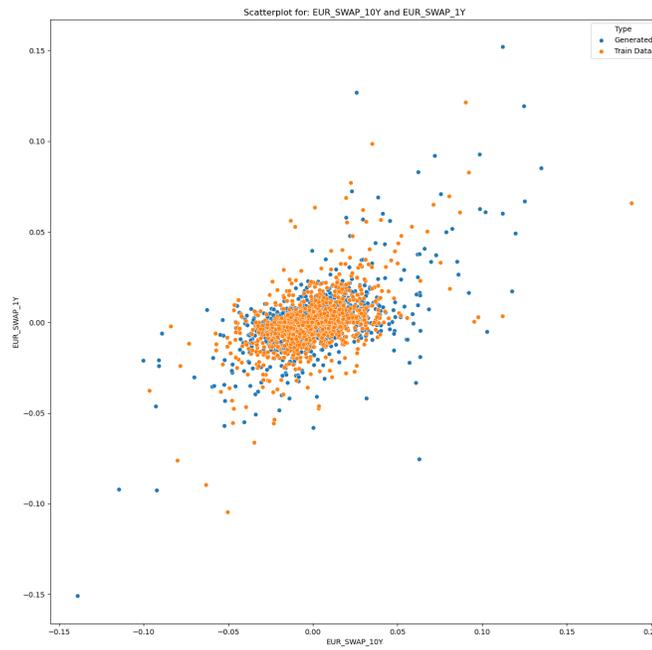


Marginalverteilung: IT Gov-Spread (7Y)

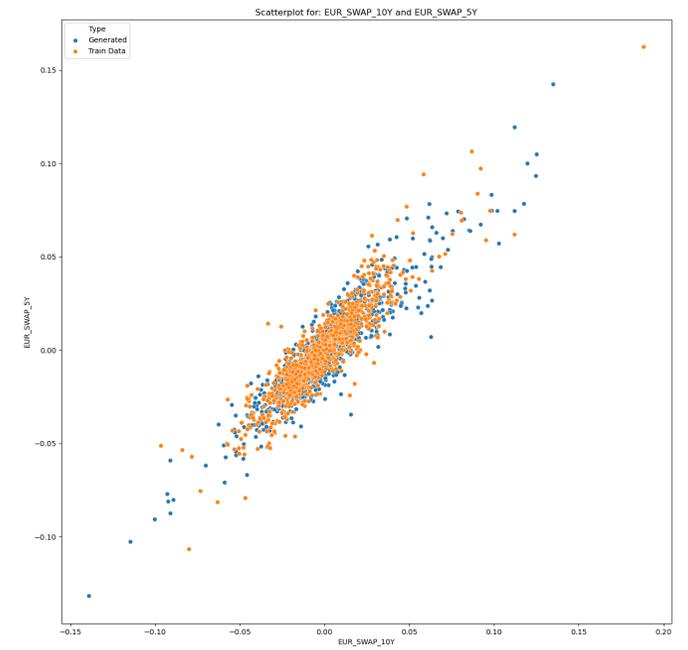
ABHÄNGIGKEITSSTRUKTUR



EUR_SWAP_1Y vs. EUR_SWAP_5Y

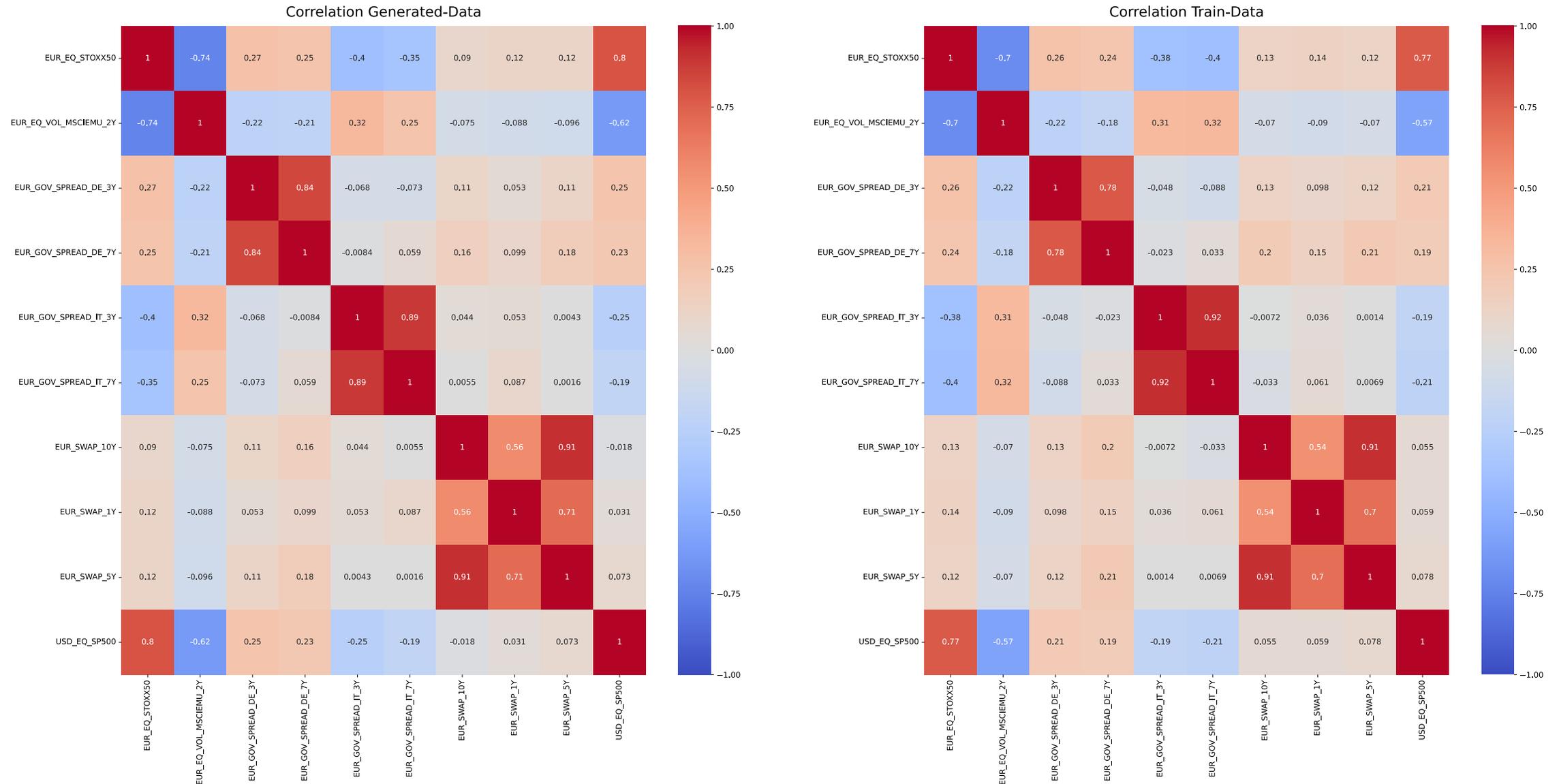


EUR_SWAP_10Y vs. EUR_SWAP_1Y



EUR_SWAP_10Y vs. EUR_SWAP_5Y

VERGLEICH DER ABHÄNGIGKEITSSTRUKTUR



FALLSTUDIE 4: ANNUALISIERUNG

Idee & Umsetzung

Motivation:

- Für eine Vielzahl von Anwendungen wird eine 1-Jahres Verteilung von Risikofaktoren benötigt (z.B. 1-Jahres Value-at-Risk für Solvency II)
- Problem: Zu wenig Daten auf 1-Jahresgranularität

➤ Annualisierung:

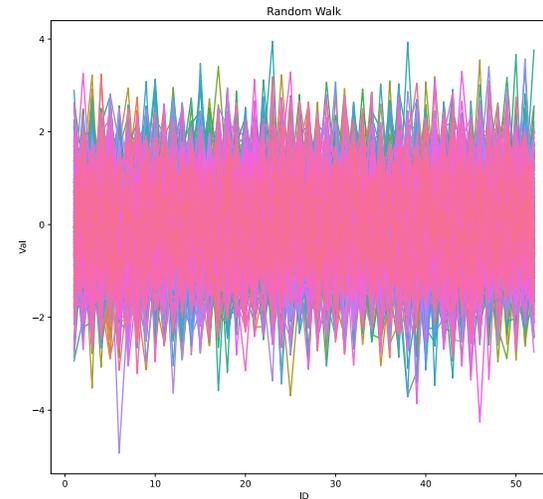
- Kalbrierung von Verteilungen auf höherer Frequenz (z.B. wöchentlich, monatlich)
- Aggregation zu niedrigerer Frequenz (jährlich) unter Berücksichtigung von Autokorrelation

➤ Gängige Verfahren:

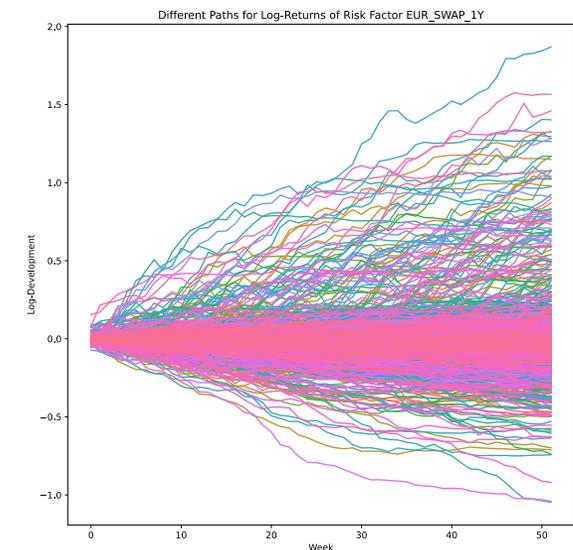
- Skalierung z.B. \sqrt{t} -Regel
 - Aggregation über Kopulas (kalibriert an historische Autokorrelation) z.B. Gauss-Kopula, T-Kopula
 - Blockbootstrap
 - Zeitreihenmodelle AR, ARMA, GARCH
- Viel manuelle Interaktion / Expertjudgement

Beispiel:

- Training von GANs zur Generierung von Zeitreihen von wöchentlichen (displaced-) log-returns
- Simulation von 1-Jahres log-returns

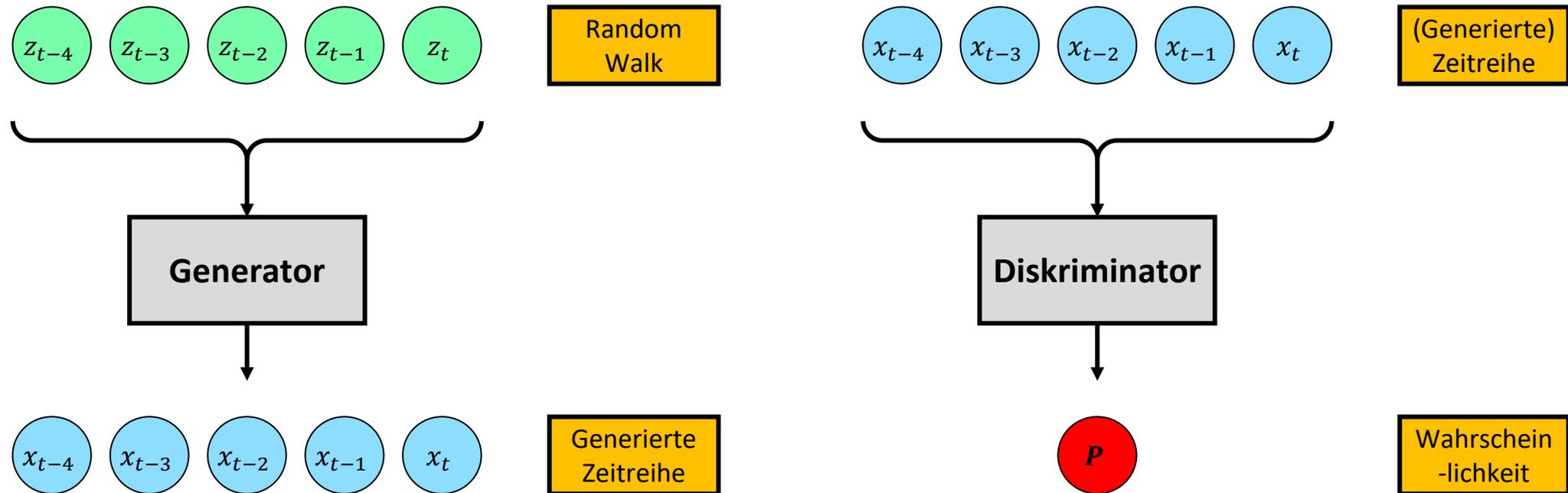


Random Walk



Generierte Zeitreihe

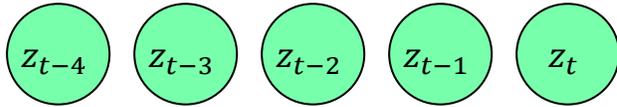
GANS ZUR GENERIERUNG VON ZEITREIHEN: IDEE



REGULAR ATTENTION LAYER

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{k=1}^n e^{z_k}} \text{ für } z \in \mathbb{R}^n$$

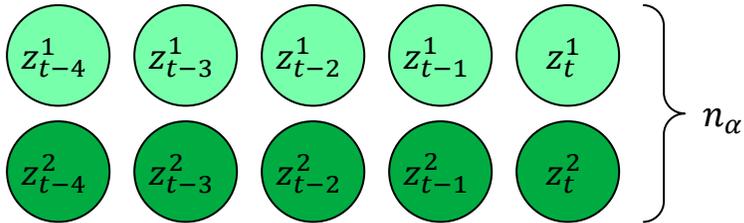
1. Input: $Z \in \mathbb{R}^{n_i \times n_i}$ (n_i Länge der Inputsequenz und n_i Dimension)



2. Einbettung in hidden dimension n_α

$$\begin{aligned} Q &= ZW^Q + 1\mathbf{b}^{Q^T} \\ K &= ZW^K + 1\mathbf{b}^{K^T} \\ V &= ZW^V + 1\mathbf{b}^{V^T} \end{aligned}$$

$W^Q, W^K, W^V \in \mathbb{R}^{n_i \times n_\alpha}$ und $\mathbf{b}^Q, \mathbf{b}^K, \mathbf{b}^V \in \mathbb{R}^{n_\alpha}$



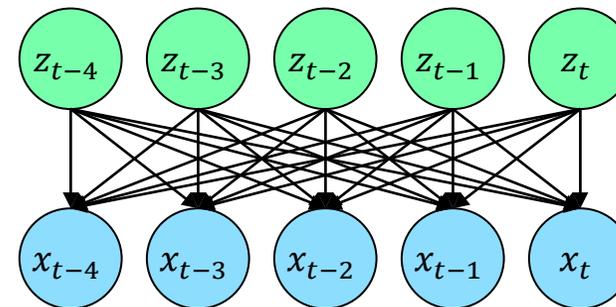
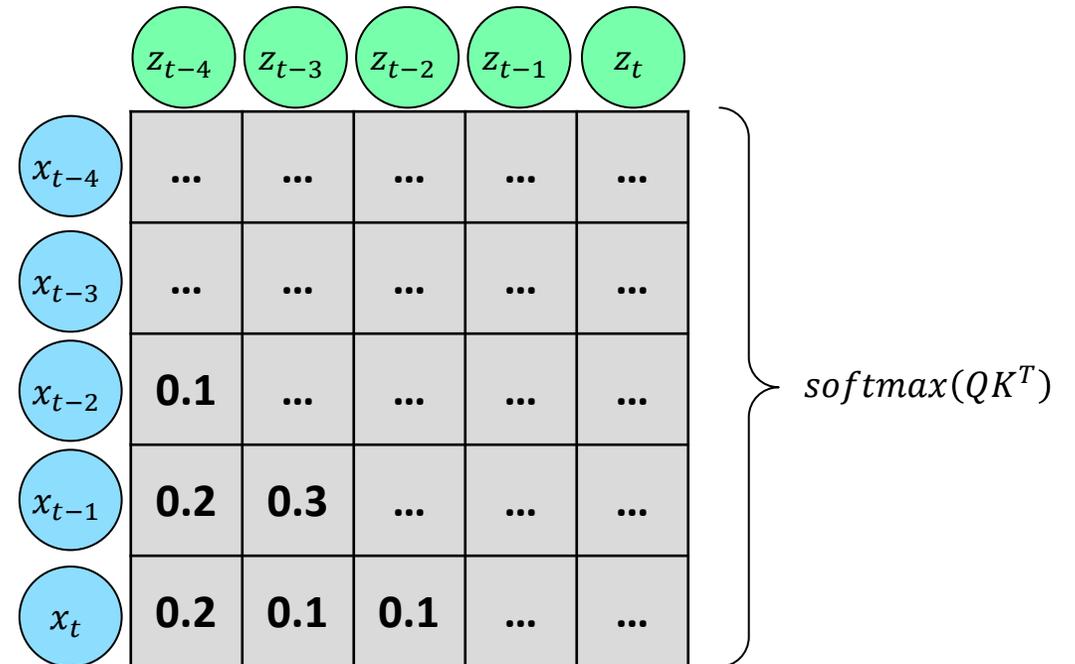
3. Berechnung von Gewichten:

$$A = \text{softmax}(QK^T)V$$

4. Output:

$$O = AW^O + 1\mathbf{b}^{O^T}$$

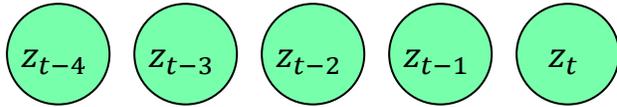
$W^O \in \mathbb{R}^{n_\alpha \times n_i}$ und $\mathbf{b}^O \in \mathbb{R}^{n_\alpha}$



CAUSAL ATTENTION LAYER

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{k=1}^n e^{z_k}} \text{ für } z \in \mathbb{R}^n$$

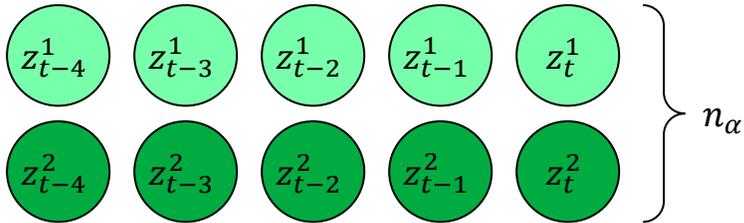
1. Input: $Z \in \mathbb{R}^{n_i \times n_i}$ (n_i Länge der Inputsequenz und n_i Dimension)



2. Einbettung in hidden dimension n_α

$$\begin{aligned} Q &= ZW^Q + 1\mathbf{b}^{Q^T} \\ K &= ZW^K + 1\mathbf{b}^{K^T} \\ V &= ZW^V + 1\mathbf{b}^{V^T} \end{aligned}$$

$W^Q, W^K, W^V \in \mathbb{R}^{n_i \times n_\alpha}$ und $\mathbf{b}^Q, \mathbf{b}^K, \mathbf{b}^V \in \mathbb{R}^{n_\alpha}$



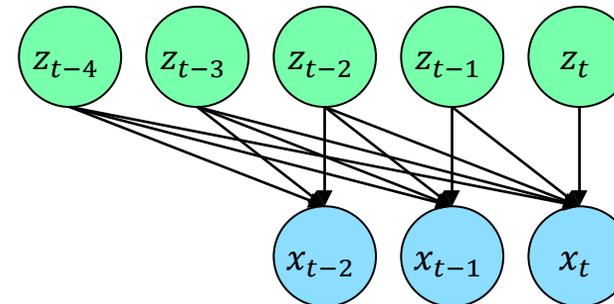
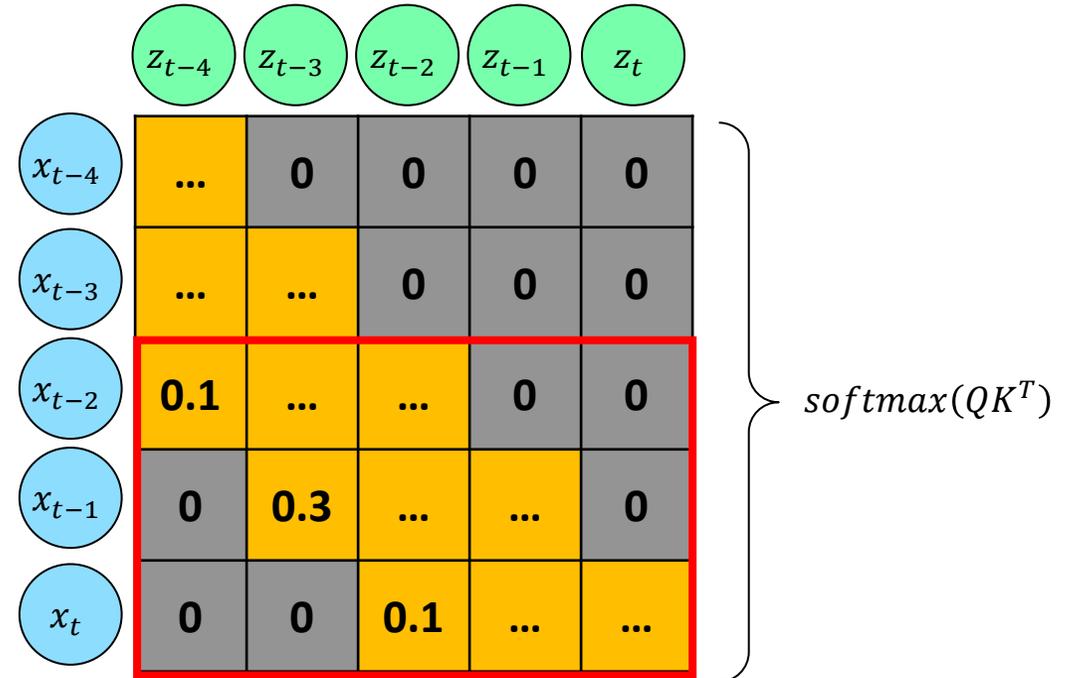
3. Berechnung von Gewichten:

$$A = \text{softmax}(QK^T + M)V$$

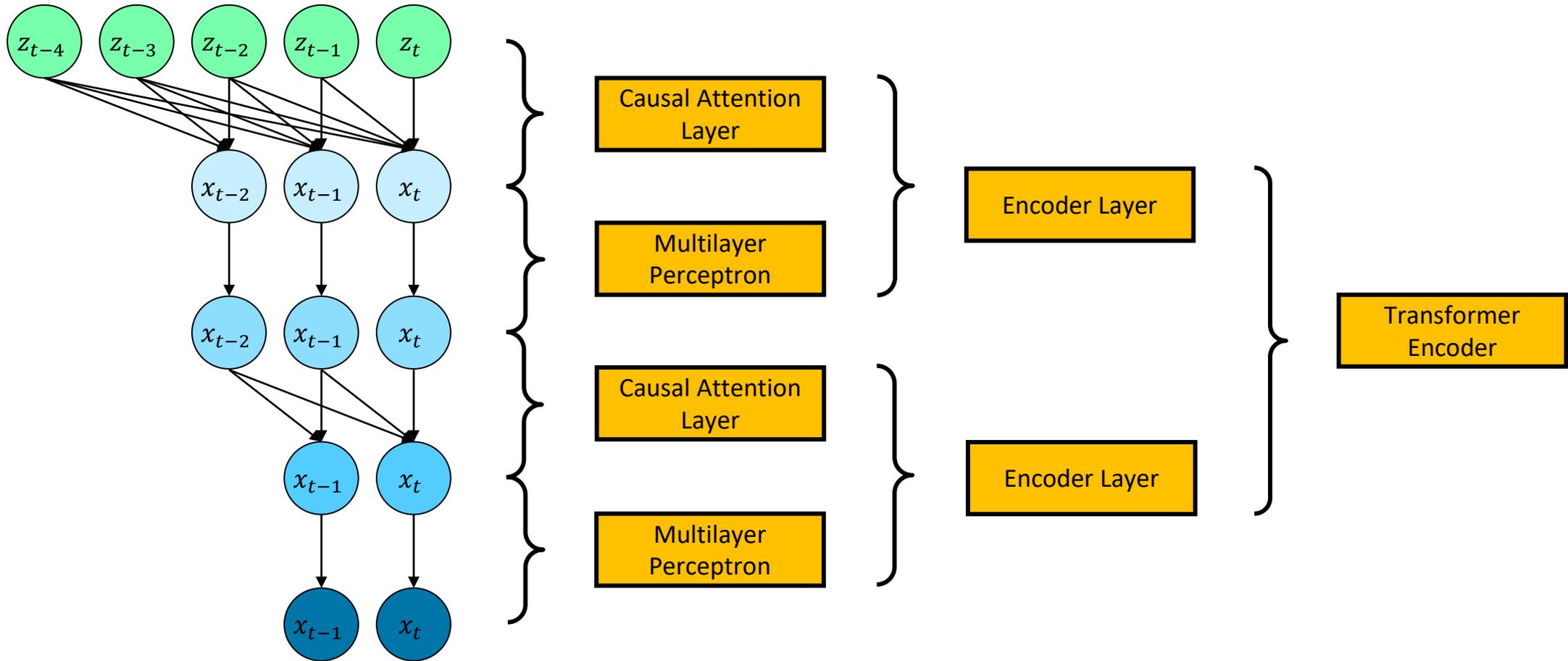
4. Output:

$$O = AW^O + 1\mathbf{b}^{O^T}$$

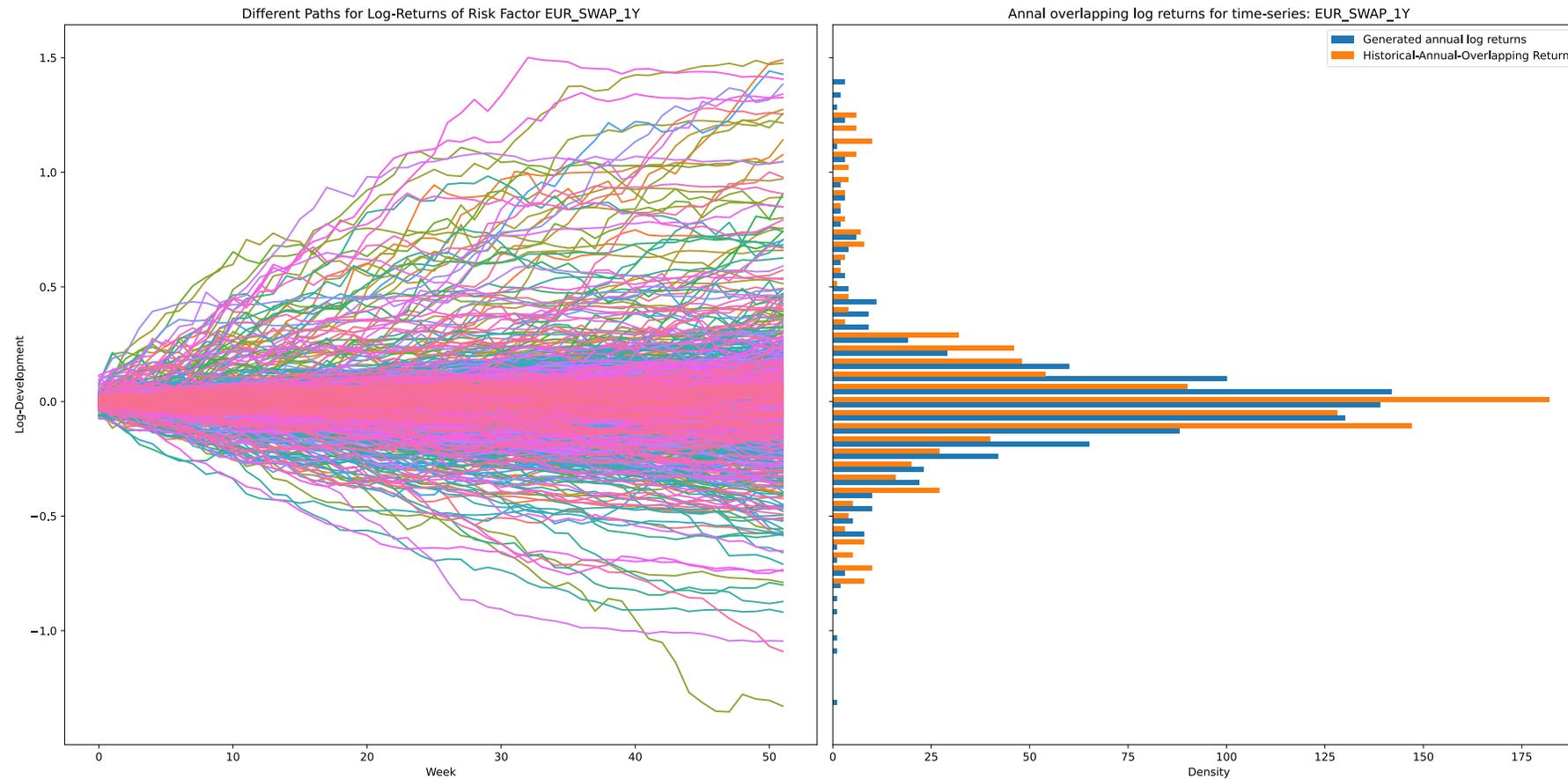
$W^O \in \mathbb{R}^{n_\alpha \times n_i}$ und $\mathbf{b}^O \in \mathbb{R}^{n_\alpha}$



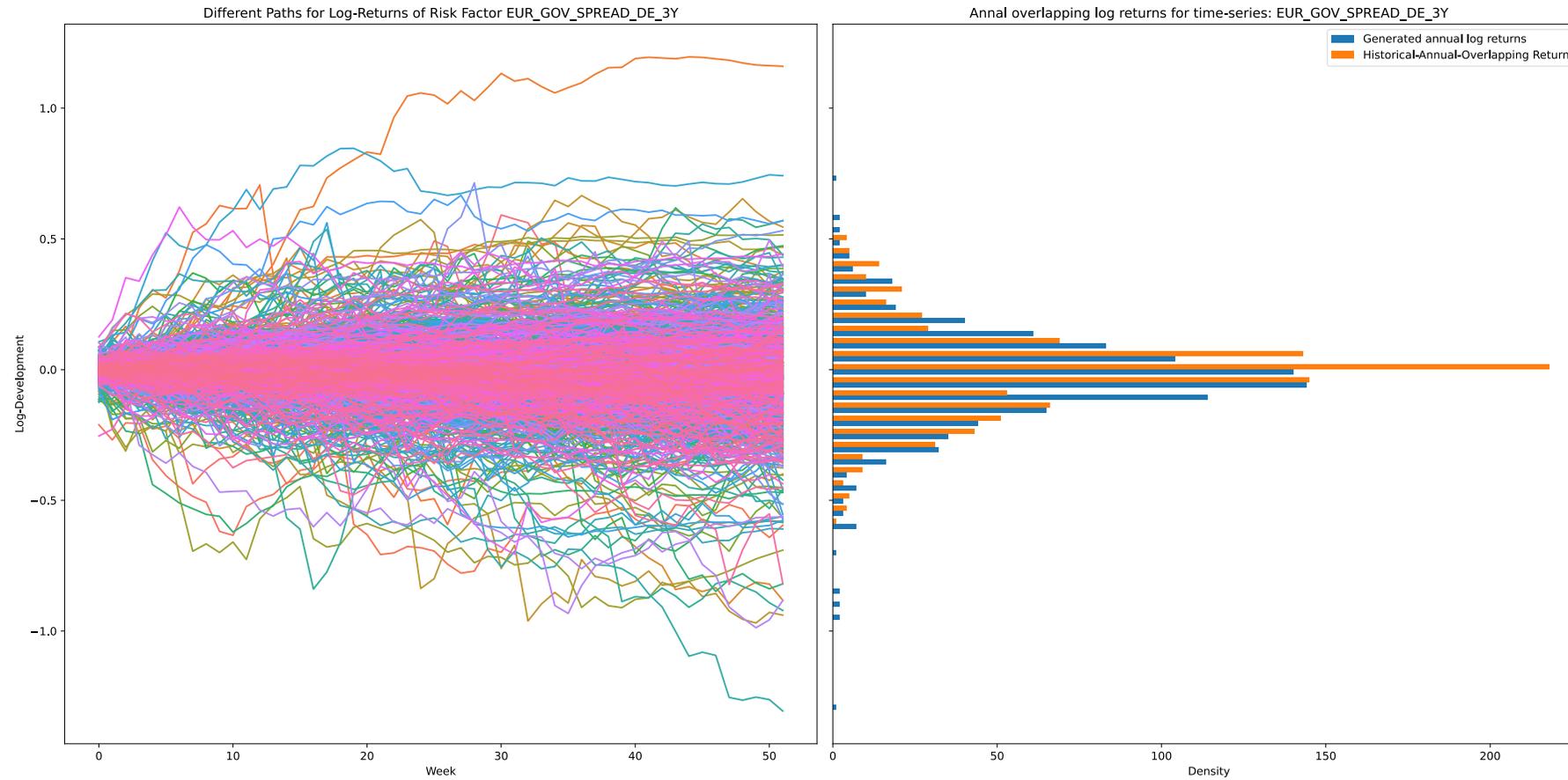
TEMPORAL TRANSFORMER GAN



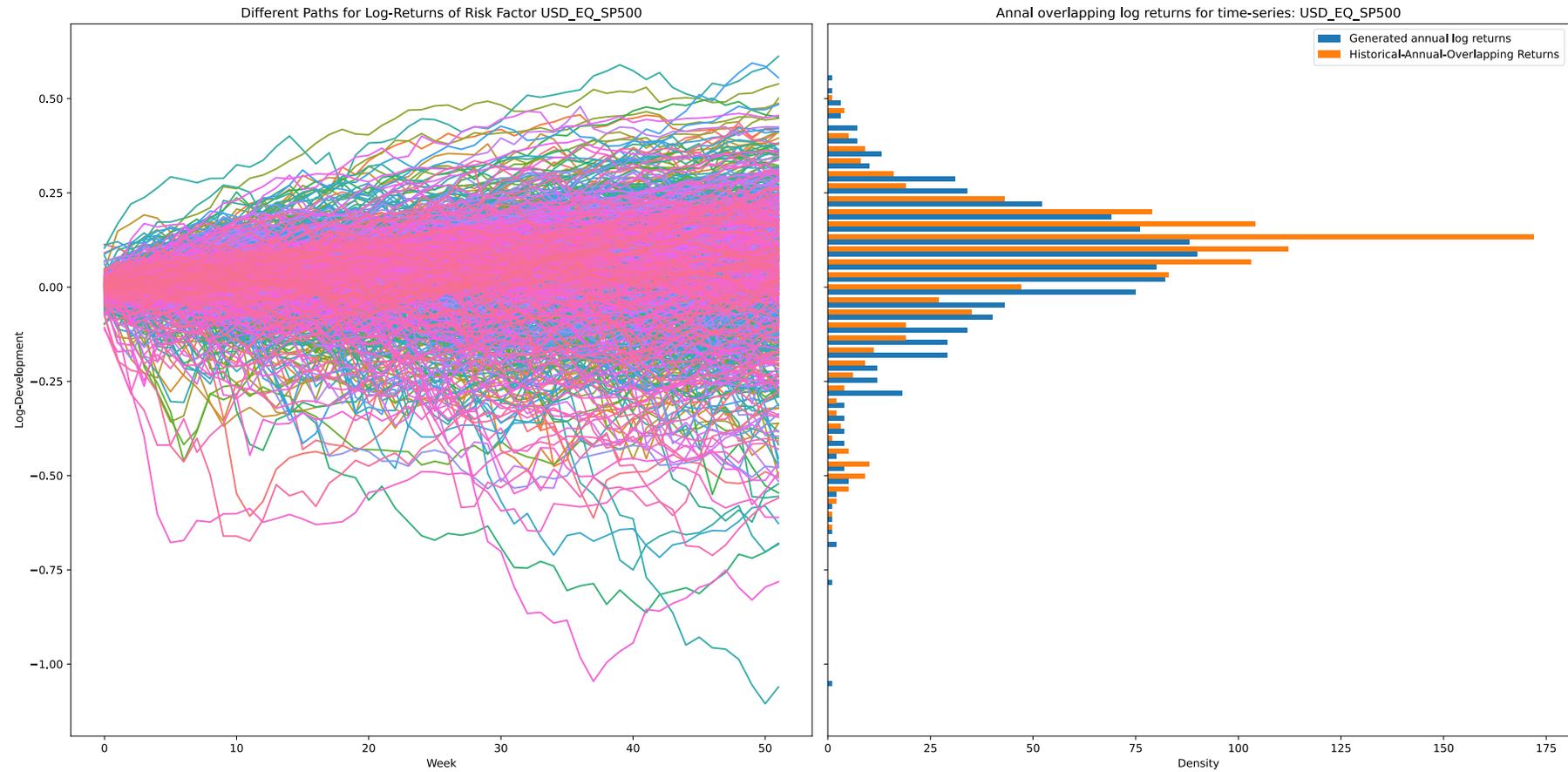
ERGEBNISSE FÜR EUR – SWAPRATE (1Y)



ERGEBNISSE FÜR DE GOV-SPREAD (3Y)



ERGEBNISSE FÜR EQ-SP500



WRAPPING UP ...

UNSER FAZIT

Wir denken, dass generative KI großes Potential besitzt, unseren Arbeitsalltag zu erleichtern, bspw. in Bereichen wie ...

- **Content-Erstellung:** Generative KI kann bei der Erstellung von Texten, Artikeln, Berichten und anderen Inhalten unterstützen. Sie kann dabei helfen, Ideen zu generieren, Texte zu verfassen und kann dabei sogar verschiedene Schreibstile verwenden (formlose E-Mail, technische Dokumentation, ...).
- **Kundenservice:** Generative KI kann in Chatbots und virtuellen Assistenten eingesetzt werden, um Kundenanfragen zu beantworten und Probleme zu lösen. Sie kann dabei helfen, häufig gestellte Fragen zu beantworten und personalisierte Unterstützung zu bieten.
- **Datenanalyse und Prognosen:** Generative KI kann große Mengen an Daten analysieren und Muster erkennen, um Prognosen und Risikovorhersagen zu treffen. Damit kann sie helfen, komplexe Daten zu verstehen und fundierte Entscheidungen zu treffen.
- **Automatisierung von Aufgaben:** Generative KI kann repetitive Aufgaben automatisieren und so die Effizienz und Produktivität steigern. Sie kann dabei helfen, Zeit und Ressourcen zu sparen und menschliche Mitarbeiter von monotonen Aufgaben zu entlasten.
- **Sprach- und Textverarbeitung:** Generative KI kann bei der Sprach- und Textverarbeitung eingesetzt werden, um bei Übersetzungen, Zusammenfassungen, Korrekturen und anderen sprachlichen Aufgaben zu unterstützen.



Vorsicht ist bei der Verwendung von GenAI-Modellen trotzdem geboten, denn diese fördern Risiken in den Bereichen ...

- Fehlinformation und Misuse
- Bias und Fairness
- Input Sensitivität
- Knowledge Grounding
- Erklärbarkeit und Transparenz
- Präzision, Qualität und Zuverlässigkeit
- Datenschutz



Deshalb ist es unumgänglich, gewisse Standards und Vorsichtsmaßnahmen bei der Verwendung von GenAI-Modellen einzuhalten, wie bspw. ...

- Prompt Engineering
- Data Augmentation
- Human-in-the-Loop
- Knowledge und Training

VIELEN DANK



Dr. Daniel Hohmann
Principal | Actuarial Services



Breite Strasse 28 | 40213 Düsseldorf
Mobile: +49 152 08879912
daniel.hohmann@oliverwyman.com
www.oliverwyman.com



Amelie Morgenstern
Senior Consultant | Actuarial Services



Müllerstrasse 3 | 80469 Munich
Mobile: +49 172 1365331
amelie.morgenstern@oliverwyman.com
www.oliverwyman.com



Tom Huber
Consultant | Actuarial Services



Müllerstrasse 3 | 80469 Munich
Mobile: +49 152 08879912
tom.huber@oliverwyman.com
www.oliverwyman.com



A business of Marsh McLennan