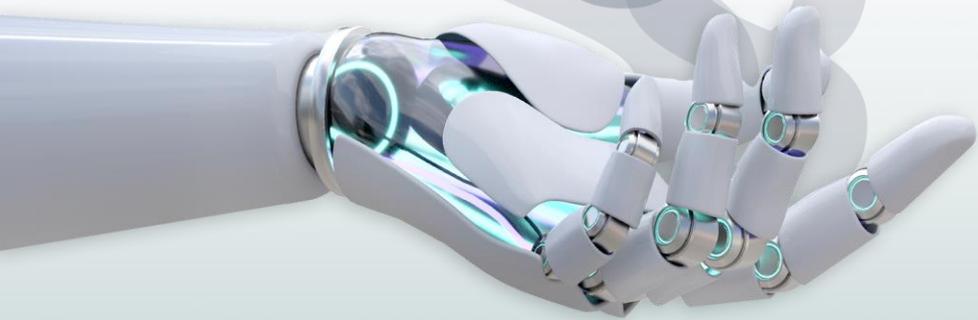


READY

FOR THE FUTURE



IDSt-TaxTechToks

Kompakt, clever, innovativ



Natural Language Processing im Steuerbereich

Prof. Dr. Rainer Bräutigam, StB

Duale Hochschule Baden-Württemberg (DHBW, Standort Mosbach), 25.07.2024

Moderator

Dietmar Gegusch
Deloitte



- Steuerpolitik
- Learning und Knowledge Management
- Verantwortlich für IDSt-Akademie
- Stellv. Vorsitzender FA VIII

Referent

Prof. Dr. Rainer Bräutigam, StB
DHBW Mosbach



- Mehrjährige Tätigkeit als Steuerberater
- Tätigkeit bei Taxy.io bis 11/2023
- Professor für Betriebswirtschaftliche Steuerlehre und Unternehmensrechnung an der DHBW
- Lehre: Unternehmensbesteuerung (u.a. auch Einsatz von KI/Daten)
- Forschung: KI im Steuerbereich (u.a.)



Agenda

1. Einführung in Natural Language Processing (NLP)
2. NLP Modelle im Steuerbereich
3. „Praxisleitfaden“

1. Einführung in Natural Language Processing (NLP)

Überblick



Information Retrieval
Question Answering

Suche
Direktantworten in Google

...

Machine Translation
Conversational AI

DeepL
ChatGPT

...

Information Retrieval ist ein Bereich der Informatik, der sich mit dem Auffinden und Bereitstellen von Informationen aus großen Datenmengen, hauptsächlich Texten, beschäftigt. Es umfasst Suchvorgänge, Indizierung und Datenabfragen zur Identifikation relevanter Informationen.

1. Einführung in Natural Language Processing (NLP)

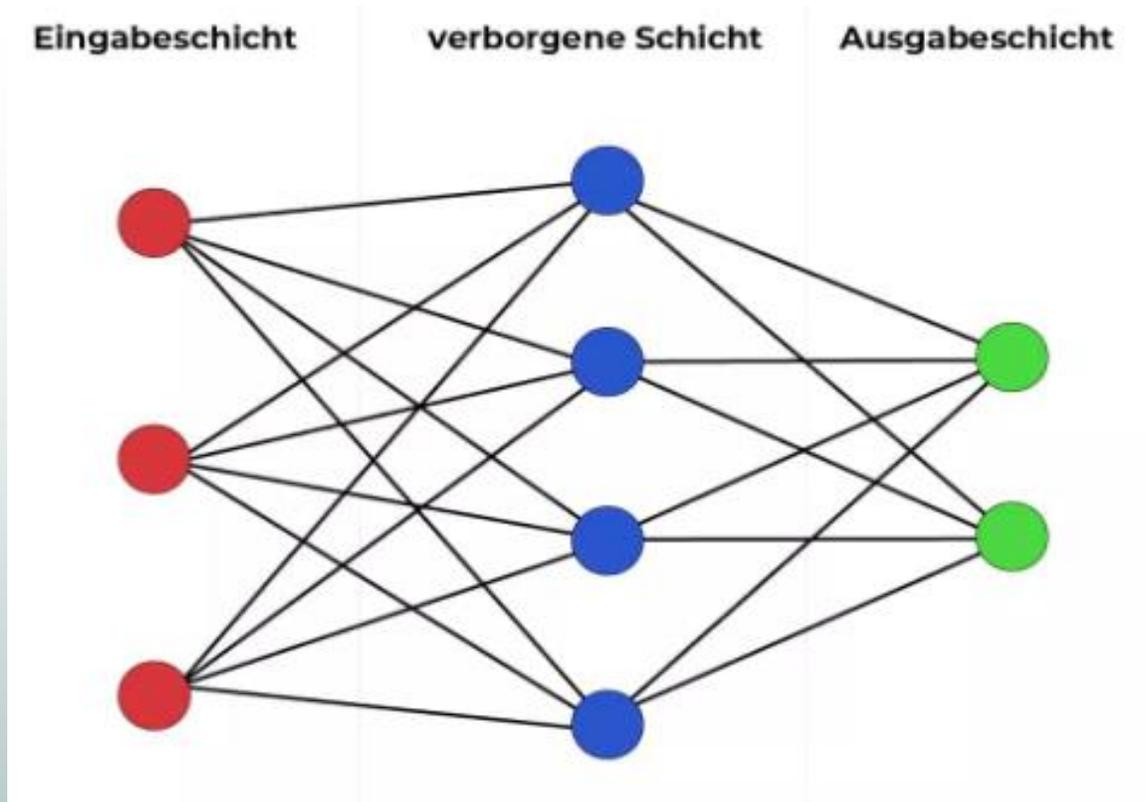
Funktionsweise

- Neuronale Netze sind Algorithmen, die dem menschlichen Gehirn nachempfunden sind.
- Ein “tiefes” neuronales Netz zeichnet sich durch eine mehrschichtige Architektur aus, bei der zwischen „Eingabe- und Ausgabeschicht“ verschiedene weitere Schritte erfolgen, die dem Nutzer mitunter verborgen bleiben.
- Grundprinzip ist die Idee der „Next Token Generation“:
 - Anhand der Nutzeranfrage wird durch mehrstufige Berechnungen die „wahrscheinlichste“ Anfangssilbe/-wort gesucht.
 - Ausgehend von der Nutzeranfrage und der ersten Anfangssilbe/-wort wird die wahrscheinlichste zweite Anfangssilbe/-wort gesucht.
 -
 - Ausgehend von der Nutzeranfrage und aller bisher gebildeten Wörter wird die wahrscheinlichste letzte Silbe gesucht.
- Nachgelagert sind Feinabstimmungen unter menschlichem Einsatz notwendig, damit die Antworten des Modells nützlich, faktisch korrekt und unbedenklich sind.

1. Einführung in Natural Language Processing (NLP)

Funktionsweise

Beispiel für „neuronales Netzwerk im Alltag“: Lichtschalter



Quelle: <https://datasolut.com/neuronale-netzwerke-einfuehrung/>, Abruf am 20.07.2024

1. Einführung in Natural Language Processing (NLP)

Funktionsweise

Beispiel für die „Next Token Prediction“:

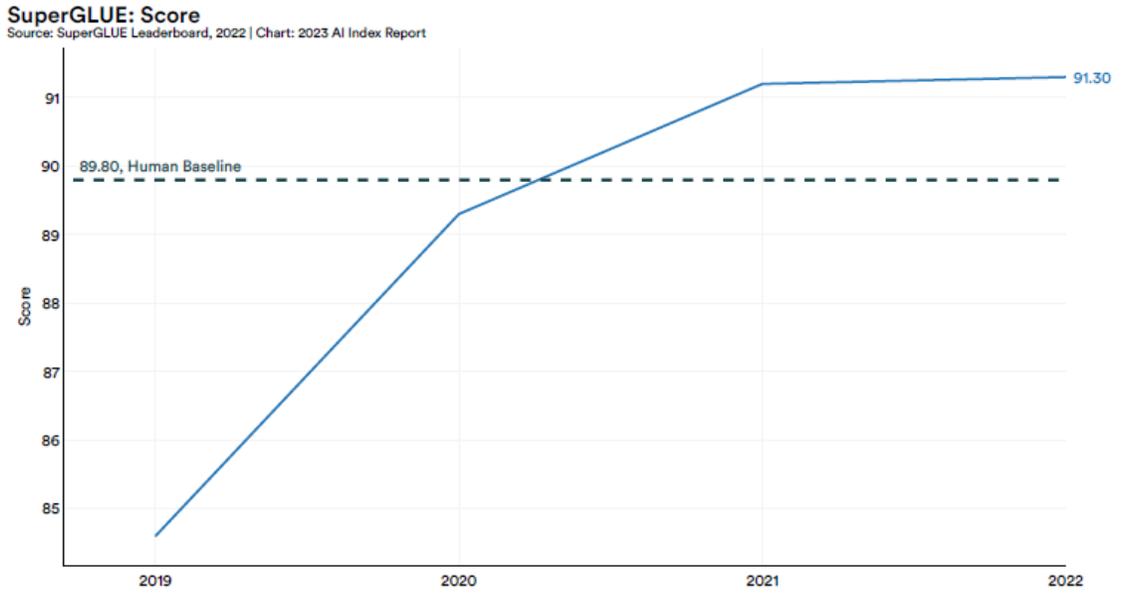
„Wie hoch ist der Körperschaftsteuersatz in Deutschland?“

Der Körperschaftsteuersatz beträgt 15% .

1. Einführung in Natural Language Processing (NLP)

„Intelligenz“ von künstlicher Intelligenz:

Standardisierte Tests zur Messung (SuperGLUE zum Leseverständnis)



Quelle: Clark et al., Artificial Intelligence Index Report 2023, S. 100

Eigener Versuch:



You

Welche der folgenden Sätze ergibt mehr Sinn?

Ich habe eine sehr gute Note bekommen, weil ich fleißig studiert habe.
Ich habe fleißig studiert, weil ich eine gute Note bekommen habe.



ChatGPT

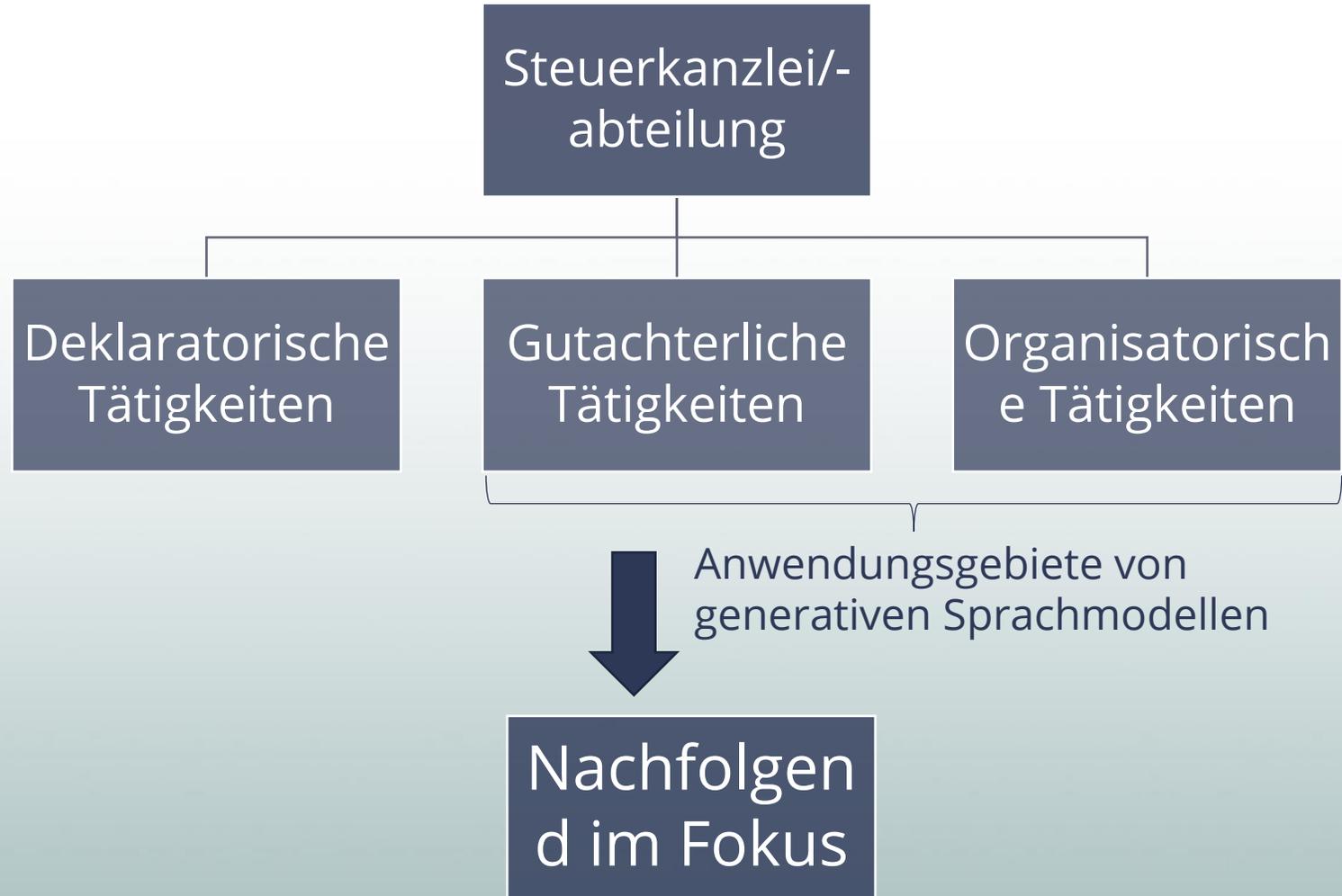
Der Satz, der mehr Sinn ergibt, ist der erste: "Ich habe eine sehr gute Note bekommen, weil ich fleißig studiert habe." Dieser Satz stellt eine kausale Beziehung her, bei der die Ursache (fleißiges Studieren) zu einer Wirkung (eine sehr gute Note bekommen) führt. Es ist logisch und üblich, dass gutes Studieren zu guten Noten führt.

Der zweite Satz, "Ich habe fleißig studiert, weil ich eine gute Note bekommen habe," stellt die Ereignisse in einer ungewöhnlichen Reihenfolge dar, die weniger logisch erscheint. Normalerweise studiert man fleißig, um eine gute Note zu bekommen, und nicht umgekehrt.

Agenda

1. Einführung in Natural Language Processing (NLP)
- 2. NLP im Steuerbereich**
3. „Praxisleitfaden“

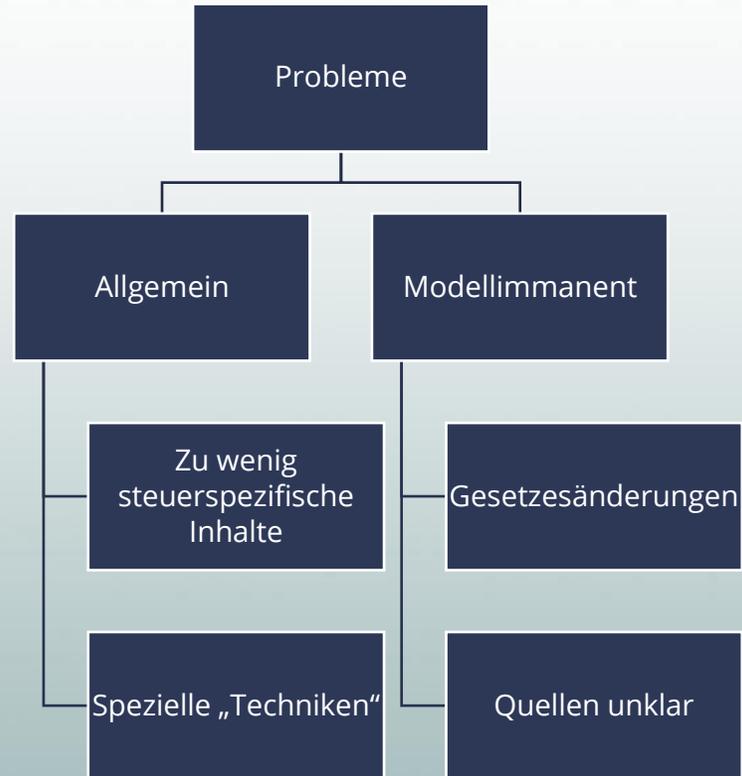
2. NLP im Steuerbereich



2. NLP im Steuerbereich

Einsatz eines allgemeinen Sprachmodells

- Prüfung von steuerfachlichen Aufgaben mit Hilfe von ChatGPT:
 - ChatGPT 3.5.: knapp nicht bestanden
 - ChatGPT 4.0: knapp bestanden
- Quelle: Volovskiy et al., Rethinking Tax 2/2023, S. 9-11.
- Probleme beim Einsatz eines allgemeinen Sprachmodells:



2. NLP im Steuerbereich

Kombination aus Sprachmodell und Information Retrieval

- Um diese „systemischen“ Schwächen abzustellen und generative Sprachmodelle auch für den steuerfachlichen Bereich nutzbar zu machen, bedarf es einer anderen Herangehensweise.
- Als zielführend hat sich eine Erweiterung um ein „Information Retrieval“ herausgestellt :
 1. Literaturdokumente werden durchsuchbar gemacht (Unterteilung in kleinere Einheiten, „Vektorisierung“).
 2. Nutzeranfrage wird ebenfalls vektorisiert.
 3. Zwischen den Fragmenten der Literaturdokumente und der Nutzeranfrage werden Ähnlichkeitswerte berechnet.
 4. Nur Textstellen, die einen hohen Ähnlichkeitswert aufweisen, werden selektiert.
 5. Die Nutzeranfrage sowie die ähnlichsten Textstellen werden dem Sprachmodell zur Beantwortung vorgelegt.
- In diesem (erweiterten) Ansatz kommt dem „Suchen der besten Textstellen“ eine sehr viel bedeutendere Funktion zu.
- Das Sprachmodell dient in diesem Fall nur noch zum „Formulieren der Antwort“ anhand der Nutzeranfrage und den gefundenen Textstellen.

2. NLP im Steuerbereich

Kombination aus Sprachmodell und Information Retrieval

Beispiel:

Leitsätze

1. (...) von einer grundbesitzverwaltenden Personengesellschaft auf einem Weihnachtsmarkt ausgeübte (...) Tätigkeit schließt (...) die erweiterte Kürzung (...) aus.
2. Unerheblich ist, ob diese originär gewerbliche Nebentätigkeit (...) nur äußerst geringfügig ist.

Tenor

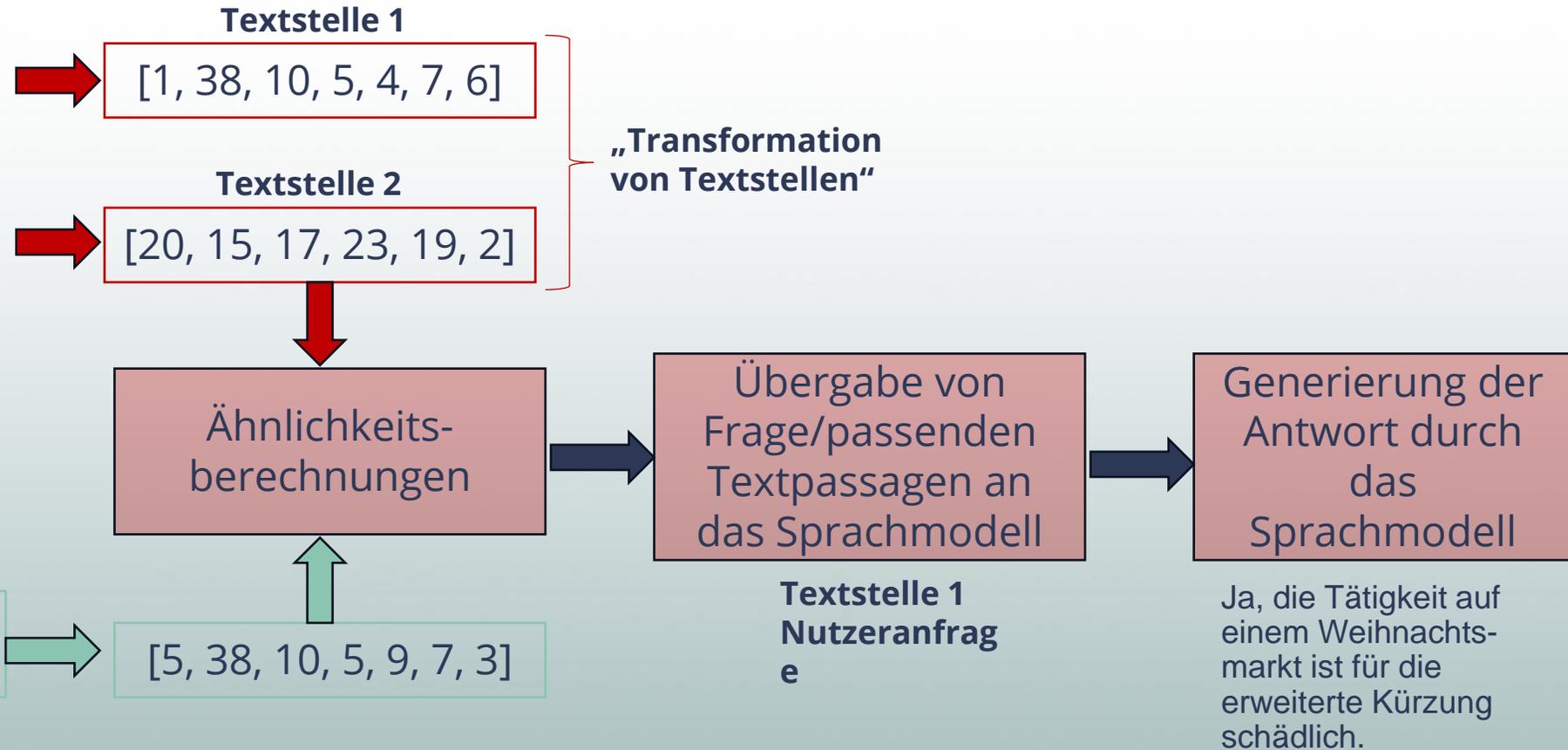
....

Tatbestand

....

Nutzeranfrage

„Ist eine Tätigkeit auf einem Weihnachtsmarkt für die erweiterte Kürzung schädlich?“



Quelle: Eigene Darstellung in Anlehnung an Diller/Ehm/Katzlmayr/Späth, beck.digitax 2023, S. 276.

Agenda

1. Einführung in Natural Language Processing (NLP)
2. NLP im Steuerbereich
3. **„Praxisleitfaden“**

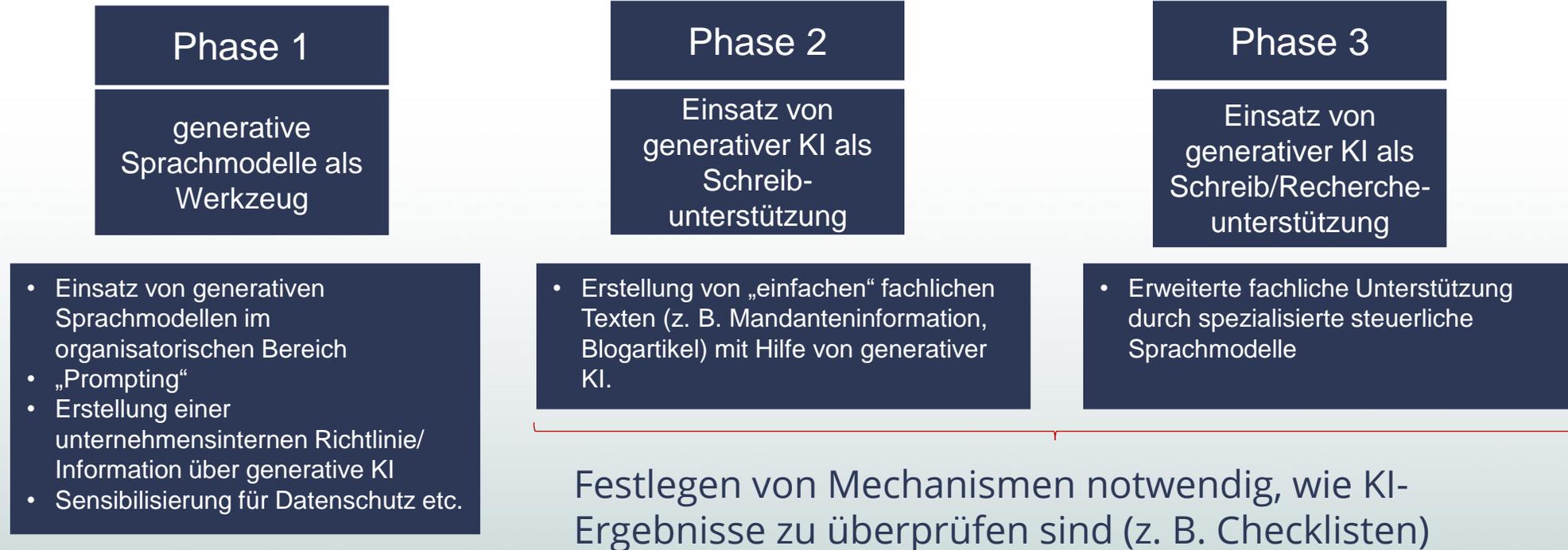
3. „Praxisleitfaden“

Welches Modell wofür?

Allgemeines Sprachmodell	„Chat with PDF“	Sprachmodell + Verlags-/Dokumentendatenbank
		Suchen relevanter Dokumente
	Suchunterstützung in einem Dokument	Suchunterstützung in einem Dokument
	Formulierungsvorschlag	Formulierungsvorschlag
Formulierungsverbesserung	Formulierungsverbesserung	Formulierungsverbesserung

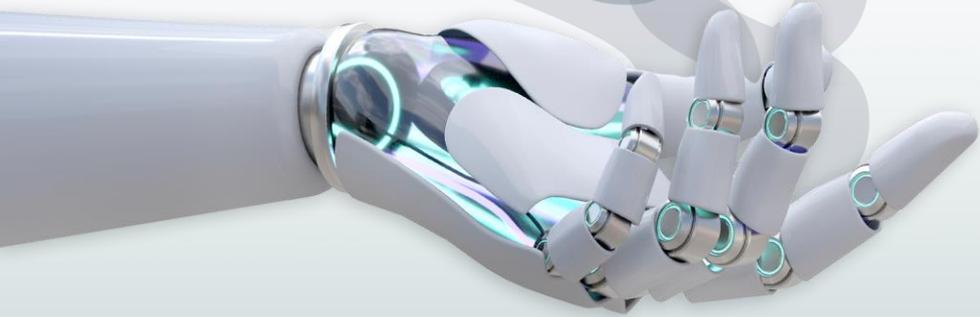
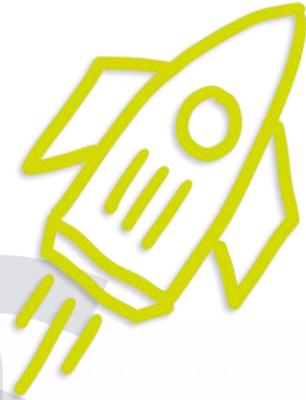
3. Praxisleitfaden

Integration in die Berufspraxis:



READY

FOR THE FUTURE



Vielen Dank!

Veranstaltungen

Donnerstag, 23.05.2024 - 12:30 Uhr bis 13:00 Uhr

Robotics Process Automation im Steuerbereich



Donnerstag, 20.06.2024 - 12:30 Uhr bis 13:00 Uhr

Process Mining für steuerliche Prozesse



Donnerstag, 25.07.2024 - 12:30 Uhr bis 13:00 Uhr

Natural Language Processing im Steuerbereich



Donnerstag, 19.09.2024 - 12:30 Uhr bis 13:00 Uhr

Schnittstellen / API



Donnerstag, 24.10.2024 - 12:30 Uhr bis 13:00 Uhr

Low Code / No Code für steuerliche Anwendungen



Donnerstag, 21.11.2024 - 12:30 Uhr bis 13:00 Uhr

Blockchain im Besteuerungsprozess



Donnerstag, 19.12.2024 - 12:30 Uhr bis 13:00 Uhr

Cloud (Computing) im Steuerbereich



Weitere kostenfreie Webcasts:

