

ki studie
2024

Künstliche Intelligenz im Steuerbereich

Innovationsstudie zum Potenzial und
zur technologischen Entwicklung

Mitwirkende

Projektleiter

Prof. Dr. Peter Fettke

DFKI GmbH
Campus D 3 2
66123 Saarbrücken

+49 (0) 681 85775 5142
peter.fettke@dfki.de

Michel Braun

WTS Group AG
Steuerberatungsgesellschaft
Klaus-Bungert-Straße 7
40468 Düsseldorf

Telefon +49 211 20050942
Michel.Braun@wts.de

Autoren

Alexander Beuther
Alexander Rombach
Sebastian Stephan
Prof. Dr. Peter Fettke
Dr. Jenny Köppe-Karkutsch
Martin Dönnebrink

Unternehmensvertreter

Unser ganz besonderer Dank gilt den hier genannten Unternehmensvertretern und deren Kollegen und Mitarbeitern, welche die Studie unterstützt haben.

Mitwirkende

Weiterhin bedanken wir uns bei den WTS-Mitwirkenden:

Robert Welzel
Andreas Riedl
Lars Bax
Christiane Noatsch
Dr. Philipp Besson
Gabriele Heemann
Nils Bleckmann
Martin Loibl
Ulrike Schellert
Alexandra Klein
Holger Bauer
Dirk Beduhn
Dominik Lipp
Jochen Würges
Christiane Belz
Ellen Birkemeyer
Matthias Wulf
Brigitte Fischer
Dr. Hans Maier
Michael Kortenhaus
Michael Eichelbeck
Jonah Boot
Valentin Schmidbaur

VORWORT

Liebe Leserin, lieber Leser,

der Einfluss des ChatGPT-Effekts macht sich auch im Bereich der Steuerwelt bemerkbar. Die Entstehung neuer Schlüsseltechnologien eröffnet unserer Branche vollkommen neue Perspektiven. Zugleich wird der Zugang zu KI-Anwendungen zunehmend einfacher, und ihre Nutzung erfordert oft keine ausgeprägten technischen Vorkenntnisse mehr. **Wir stehen zweifellos an der Schwelle einer Zeitenwende!**

Im Jahr 2016 haben wir gemeinsam mit dem DFKI unsere erste Studie "Künstliche Intelligenz im Steuerbereich" veröffentlicht. Die damals gewonnenen Erkenntnisse waren äußerst beeindruckend und verdeutlichten einmal mehr das vielversprechende Zusammenspiel von Steuern und KI. Im Steuerbereich werden große Datenmengen verarbeitet und zahlreiche wiederkehrende sowie manuelle Tätigkeiten ausgeführt. Daraus ergibt sich ein enormes Potenzial für Automatisierung – und genau das haben wir seit 2016 erlebt.

In den letzten Jahren wurden steuerliche Prozesse in nahezu allen Steuerdisziplinen zunehmend automatisiert, wobei in vielen Anwendungen bereits KI zum Einsatz kommt. Doch was wir seit 2022 erleben, wird unsere Arbeitsweise für immer verändern. Viele der KI-Basistechnologien haben mittlerweile einen hohen technischen Reifegrad erreicht, was zu einem enormen Schub bei KI-Anwendungen geführt hat. Heute geht es nicht nur um die Automatisierung von Prozessen, sondern auch um wissensbasierte Aufgaben, die von KI-Anwendungen mit beeindruckender Genauigkeit erledigt werden können. Dies macht sie zu idealen digitalen Assistenten.

Im Zentrum stehen generative Sprachmodelle, die aufgrund ihrer Fähigkeit zur Sprachverarbeitung eine entscheidende Rolle spielen. Zudem wurden im Laufe der Zeit viele weitere KI-Verfahren entwickelt, die zukünftig zur Unterstützung von Aufgaben in Steuerfunktionen eingesetzt werden können. Es ist also höchste Zeit, in eine weitere Runde zu gehen und den aktuellen Stand in unserer mittlerweile dritten Auflage der Studie genauer zu untersuchen.

Vorweggenommen sei:

Die Ergebnisse sind äußerst spannend und zeigen im Vergleich zu den Vorgängerstudien aus den Jahren 2016 und 2017 einen erheblichen Entwicklungssprung bei KI-Anwendungen auf. Wir stehen kurz vor einem Durchbruch, durch den sich KI-Anwendungen schon sehr bald als Standardtechnologie in Steuerabteilungen etablieren werden.

Ich wünsche Ihnen viel Freude beim Lesen.

Ihr Fritz Esterer
CEO der WTS-Gruppe

Fritz Esterer
CEO der WTS-Gruppe



Inhalt

Abbildungsverzeichnis	6
Abkürzungsverzeichnis	7
Zusammenfassung der KI-Studie	8
1 Einleitung	9
1.1 Ausgangslage	9
1.2 Motivation	10
1.3 Kernthesen	11
1.4 Zielsetzung	12
1.5 Methode	13
1.6 Aufbau und Struktur	14
2 Aktuelle Entwicklungen der Künstlichen Intelligenz	15
2.1 Einführung	15
2.2 Maschinelles Lernen	16
2.3 Große Sprachmodelle	18
2.4 Intelligente Robotergestützte Prozessautomatisierung	20
2.5 Erklärbare Künstliche Intelligenz	22
2.6 Anomalieerkennung	24
2.7 Klassifikationsmethoden	26
2.8 Synthetische Daten	27
3 KI-Anwendungen im Steuerbereich – Ergebnisse der Experteninterviews	29
3.1 Einführung	29
3.2 Anwendungen im Bereich HR Taxes	30
3.3 Anwendungen im Bereich Umsatzsteuer	34
3.4 Anwendungen im Bereich Ertragsteuer	37
3.5 Anwendungen im Bereich Zoll	41
3.6 Anwendungen im Bereich Verrechnungspreise	44
3.7 Anwendungen im Bereich Tax CMS / steuerliches Risikomanagement	47
3.8 Anwendungen im Bereich International Tax	50
3.9 Anwendungen im Bereich Change Management	53
3.10 Anwendungen im Bereich FS Tax	56
3.11 Anwendungen im Bereich Tax Allgemein - Of Counsel Hans Maier	59
3.12 Anwendungen im Bereich Tax Allgemein - Of Counsel Brigitte Fischer	61

4	Anwendungsszenarien	64
4.1	Umsatzsteuerliche Klassifizierung von Leistungsbeschreibungen von Rechnungen	64
4.2	Automatisierte Generierung von Verrechnungspreisdokumentationen	67
4.3	Informationsextraktion aus Grundsteuerbescheiden	70
4.4	Automatisierung des Transfer Pricing Benchmarking Prozess	74
4.5	Synthetische Daten für die Anomalieerkennung	78
4.6	Kombination großer Sprachmodelle mit Wissensgraphen	81
5	Schlussfolgerung	85
	Glossar	86
	Literatur	91

Abbildungsverzeichnis

Abbildung 1:	Entwicklung der Künstlichen Intelligenz	15
Abbildung 2:	Übersicht Maschinelles Lernen [1].....	17
Abbildung 3:	Überblick über iRPA	20
Abbildung 4:	Evaluationsergebnisse aller Prompts für <i>GPT-4</i> und <i>LeoLM</i>	66
Abbildung 5:	Lösungsarchitektur des Assistenzsystems	69
Abbildung 6:	Konzeption der Informationsextraktion	71
Abbildung 7:	Benutzeroberfläche zur Datenvalidierung	72
Abbildung 8:	Transfer Pricing Benchmarking Prozess	75
Abbildung 9:	User Interface des TP AI-Benchmarking Tools.....	76
Abbildung 10:	Kennzahlen zur Performance der Textklassifikation	77
Abbildung 11:	Vorgehensweise der Generierung synthetischer Anomalien	79
Abbildung 12:	Beispiel eines Wissensgraphen für den steuerrechtlichen Rahmen.	82
Abbildung 13:	Kombination eines großen Sprachmodells mit einem Knowledge Graph [61]	83
Abbildung 14:	Vor- und Nachteile von großen Sprachmodellen und Knowledge Graphs [61].....	83
Abbildung 15:	Antworten aus klassischem RAG-System, wts plAIground.....	84

Abkürzungsverzeichnis

BEPS	Base Erosion and Profit Shifting
BERT	Bidirectional Encoder Representations from Transformers
CbC	Country-by-Country
CTGAN	Conditional Tabular Generative Adversarial Network
DSGVO	Datenschutz-Grundverordnung
GAN	Generative Adversarial Network
GPT	Generative Pre-trained Transformer
GuV	Gewinn- und Verlustrechnung
ICL	In-Context Learning
iRPA	Intelligente Robotergestützte Prozessautomatisierung
KI	Künstliche Intelligenz
LIME	Local Interpretable Model-agnostic Explanations
LLM	Large Language Model
LSTM	Long Short-Term Memory
ML	Maschinelles Lernen
MLOps	Machine Learning Operations
NLP	Natural Language Processing
OCR	Optical Character Recognition
PEFT	Parameter-Efficient Fine-Tuning
RNN	Recurrent Neural Network
RAG	Retrieval-Augmented Generation
RPA	Robotic Process Automation
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
XAI	eXplainable AI
XGBoost	Extreme Gradient Boosting

Zusammenfassung der KI-Studie

KI-Anwendungen sind momentan in aller Munde. Doch was sind eigentlich die Gründe dafür, dass an intelligenten Software-Lösungen und modernen Technologien wie ChatGPT von OpenAI innerhalb kürzester Zeit kaum noch jemand vorbeikommt?

In erster Linie dürfte es am hohen Potenzial von Künstlicher Intelligenz (KI) liegen, mit deren Hilfe sich große Informationsmengen innerhalb kürzester Zeit auswerten und effizient verarbeiten lassen. Auch die intelligente Automatisierung kognitiv anspruchsvoller Tätigkeiten und Prozesse durch selbstlernende Algorithmen werden in Zukunft kaum mehr wegzudenken sein und in vielen Bereichen unserer Arbeitswelt für einen grundlegenden Wandel sorgen.

Schon heute haben zahlreiche Basistechnologien der KI eine hohe technische Reife erreicht und werden in den verschiedensten Anwendungsgebieten erfolgreich eingesetzt. So auch in den Bereichen der Steuerfunktion, für die sich seit der letzten KI-Innovationsstudie 2017 neue Chancen und technologische Möglichkeiten ergeben. Gerade weil für die nächsten Jahre mit einem zunehmenden Wachstum an Innovationen im KI-Bereich zu rechnen ist, gilt es die Trends von morgen schon heute genauer zu analysieren und deren Anwendungspotenzial für den Steuerbereich möglichst konkret zu bewerten.

Genau deshalb haben sich verschiedene Mandanten und Partner von WTS gemeinsam mit den Experten des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) zusammengetan, um die aktuell verfügbaren KI-Technologien zu untersuchen. Die Ergebnisse möglicher Potenziale für den Einsatz von KI sowie deren Auswirkungen auf das Tätigkeitsspektrum der Steuerfunktion werden in der vorliegenden gemeinsamen Studie der WTS und des DFKI zusammengefasst.

Auf Basis von Experteninterviews und Workshops werden dabei nicht nur wegweisende Lösungen und deren Einsatzmöglichkeiten im Steuerbereich praxisnah veranschaulicht. Auch vielversprechende und KI-basierte Anwendungsszenarien werden identifiziert, die für die Steuerfunktion von besonderer Bedeutung sind. Ferner wurden prototypische Konzepte und spezifische Anwendungsszenarien entwickelt und implementiert, um die sich durch den Einsatz von KI-Technologien ergebenden Vorteile besser aufzeigen zu können.

DIE WICHTIGSTEN ERKENNTNISSE KURZ ZUSAMMENGEFASST:

Eine Systematisierung der identifizierten Potenziale von KI hat ergeben, dass der Einsatz von Assistenzsystemen im Steuerbereich zu enormen Produktivitätssteigerungen und Qualitätsverbesserungen führen kann. Konsequenterweise werden KI-Technologien bisher gültige Prozesse maßgeblich verändern und bereits in naher Zukunft zu einem grundlegenden Wandel des Berufsbildes der Steuerberatung führen.

1 Einleitung

Die KI hat insbesondere in den vergangenen zehn Jahren deutlich an Bedeutung gewonnen und ist aus der heutigen Welt kaum mehr wegzudenken. Sei es nun im Alltag als digitaler Assistent oder im unternehmerischen Kontext als intelligenter Chatbot. Wesentliche Treiber dieser Entwicklung sind vor allem der immer einfachere Zugang zu KI-Anwendungen und deren verhältnismäßig leichte Nutzung ohne größere technische Vorkenntnisse. Nicht zu vergessen die immensen Entwicklungssprünge und Verbesserungen innerhalb der für den Einsatz benötigten IT-Infrastruktur. Einschließlich der mittlerweile bequem verfügbaren Hardwarekomponenten, der stetig steigenden Rechenleistung und der sich dadurch bedingten Kostenreduktionen.

Parallel dazu greifen immer mehr Unternehmen auf Cloud-Lösungen zurück, die einen deutlich flexibleren Zugang auf notwendige Ressourcen ermöglichen, um KI-Projekte günstig umsetzen zu können. Aber auch Daten, die im Zeitalter von Big Data und dem Internet der Dinge (IoT – Internet of Things) in einem immer größeren Umfang und in stetig besserer Qualität bereitgestellt bzw. erfasst werden können, haben zu einer immer größeren Verbreitung von KI-Lösungen beigetragen. Insbesondere das Vorhandensein sehr großer Datenmengen und die zunehmende Digitalisierung im unternehmerischen Kontext haben maßgeblichen Anteil an der aktuellen Entwicklung innovativer KI-Methoden. Denn diese benötigen in der Regel viele Trainingsdaten und vor allem Informationen in maschinenlesbarer Form, um überhaupt eine gute Performanz zu erzielen.

1.1 Ausgangslage

Die zuvor geschilderten Entwicklungen und der dadurch bedingte Rückgang an Eintrittsbarrieren haben auch im Steuerbereich dazu geführt, dass der Einsatz von KI-Systemen ein enormes Potenzial zur gezielten Effektivitäts- und Effizienzsteigerung entwickelt hat. Insbesondere im Hinblick auf die Tatsache, dass Steuerabteilungen zunehmend als eine Art „Data Hub“ im Unternehmen agieren und geprägt sind von einem umfassenden Informationsfluss. Denn gerade hier laufen sämtliche Unternehmensdaten zusammen, die dann entsprechend aufbereitet, verarbeitet und weitergeleitet werden. Weshalb der Steuerbereich aus verschiedenen Gesichtspunkten regelrecht prädestiniert ist für den Einsatz von KI-Technologien:

- › Die Steuerdomäne zeichnet sich durch ein hohes Maß an wiederkehrenden und manuellen Tätigkeiten aus. Hierunter fällt z. B. das Auslesen von Informationen aus unterschiedlichsten Dokumenten, Aufgaben im Zusammenhang mit der Buchhaltung oder die Verarbeitung von Steuererklärungen. Zur Unterstützung dieser wiederkehrenden Aufgaben können gezielt KI-Anwendungen entwickelt werden.
- › Im Steuerbereich wird mit vielen heterogenen, oft unstrukturierten Informationsquellen gearbeitet. Dazu gehören bspw. Belege oder Steuerbescheide, die verarbeitet werden müssen. KI-Methoden können hierbei eingesetzt werden, um diese komplexeren Informationsquellen automatisiert zu verarbeiten, was zuvor mithilfe bspw. regelbasierter Verfahren nur im begrenzten Umfang realisiert werden konnte.
- › Zunehmende Compliance-Anforderungen sowie sich stetig anpassende Richtlinien und Gesetze, deren Erfassung, aber auch Erfüllung für Steuerabteilungen einen erheblichen Mehraufwand bedeuten. In diesem Zusammenhang können KI-Systeme bei der Erfassung unterstützen, indem z. B. Gesetzestexte oder Urteilsprüche automatisiert analysiert und zentrale Inhalte aufbereitet werden.

Die oben genannten Punkte decken sich auch mit den Ergebnissen der bereits im Jahr 2017 gemeinsam von WTS und DFKI durchgeführten Innovationsstudie [1]. Deren Ziel war es, die Potenziale von KI im Steuerbereich genauer zu beleuchten. Im Zuge der Untersuchungen wurde festgestellt, dass der Einsatz für den Steuerbereich relevanten KI-Anwendungen zwar noch in den Kinderschuhen steckte, aber für diverse Steuerarten und den damit verbundenen Anwendungsfällen durchaus vielversprechende Möglichkeiten bestünden. Insbesondere Aufgaben der steuerrechtlichen Deklarationsberatung wurden zu jener Zeit als zentrales Anwendungsfeld identifiziert.

1.2 Motivation

Mit Blick auf die zuvor beschriebene Ausgangssituation lässt sich die Motivation der hier vorliegenden Studie wie folgt beschreiben: Sowohl aus der Forschungs- als auch der Praxisperspektive soll erfasst werden, welche Entwicklungen es seit 2017 und der ersten Studie im Zusammenhang mit KI im Steuerbereich gegeben hat. Dabei sollen insbesondere Veränderungen der Potenziale sowie gänzlich neue Möglichkeiten für den Einsatz von KI-Methoden und -Anwendungen im Steuerbereich erfasst werden.

Eine der zentralen Thesen der Studie von 2017 bezog sich dabei auf die zu erwartenden technologischen Entwicklungen. Demzufolge könne in naher Zukunft eine immer größere Anzahl an Tätigkeitsbereichen innerhalb der Steuerdomäne adressiert werden. Inwiefern sich die einst aufgestellte Behauptung im Vergleich zum aktuellen IST-Zustand bestätigen lässt, ist einer der Kernpunkte der Folgestudie von 2024. Ebenso wurden die sich aus den jüngsten Entwicklungen ergebenden Potenziale von KI-Anwendungen näher analysiert. Hierfür wurden insgesamt 11 Interviews mit Experten aus den verschiedensten Steuerbereichen wie z. B. HR Taxes, Verrechnungspreise, Change Management (vgl. ab Kapitel 3) geführt.

Auf diese Weise sollte möglichst praxisnah ermittelt werden, mit welchen Herausforderungen und Aufgaben sich der jeweilige Steuerbereich tagtäglich konfrontiert sieht und wo bereits KI-Anwendungen zum Einsatz kommen. Oder inwiefern einzelne KI-Modelle in naher Zukunft zur automatisierten Bearbeitung bestimmter Prozesse herangezogen werden können. Immer in Abhängigkeit zur jeweiligen Organisationsstruktur, den steuerlichen Fragestellungen und dem vorhandenen Potenzial für den Einsatz von KI-Methoden. So viel kann vorweggenommen werden: KI-Anwendungen wurden in der Vergangenheit bevorzugt als Analyseelement für spezifische Sachverhalte in großen Datenmengen verwendet, in der Regel zum Abbilden von klassischen Prüfroutinen. Dabei wurde allerdings oftmals noch nicht das volle Potenzial genutzt.

Die Gründe dafür waren laut Aussage der Experten neben fehlendem Budget häufig auch die mangelnde Bereitschaft, etwas zu verändern, datenschutzrechtliche Hürden und das Fehlen nötiger Skills, um KI-Anwendungen einzusetzen und damit die täglichen Herausforderungen erfolgreich zu bewältigen. Diese Grundprobleme besteht trotz der jüngsten Entwicklungen im KI-Bereich noch immer. Es wurden zwar in den letzten Jahren vergünstigte Rahmenbedingungen erschlossen, die einen KI-Einsatz in der Praxis zunehmend leichter gestalten. Ebenfalls konnten maßgebliche technologische Fortschritte erzielt werden, mit denen sich performante KI-Modelle in der Praxis integrieren lassen. Dennoch bestehen für den Steuerbereich aus praktischer Sicht nach wie vor Diskrepanzen zwischen den technologischen Möglichkeiten und den tatsächlich erschlossenen Anwendungspotenzialen bzw. Lösungsideen.

Aufgabe der Studie ist es daher, festzustellen, inwiefern diese Diskrepanzen im Verlauf der letzten Jahre ausgeräumt und vorhandene Potenziale sich durch den Einsatz von KI-Technologien nutzen lassen. Im Zuge dessen sollen auch der generelle Wandel der Steuerabteilung als solche und die notwendige Anpassung der künftigen Beratungsleistungen näher beleuchtet werden. Gerade weil die Steuerabteilungen künftig immer häufiger selbst nach einer Lösung für ein Problem suchen werden, bevor sie einen Berater konsultieren. Wobei es den Menschen als „Prüforgan“ weiterhin braucht, da selbst die fortschrittlichsten KI-Systeme der Welt Fehler machen können und noch nicht gänzlich geklärt ist, wie durch deren Einsatz ethische, moralische und haftungsrechtliche Fragen zu beantworten sind. Damit einher geht auch die Analyse des Wandels der Steuerdomäne durch KI, wie Aufgaben innerhalb des Steuerbereichs erfüllt und welche Technologien dabei künftig zum Einsatz kommen werden.

Immer mit Blick auf das mitgebrachte Potenzial der KI, ausgewählte Prozesse zu automatisieren, Steuerabteilungen zu entlasten, Kosten zu reduzieren und den Wissenstransfer zu unterstützen.

1.3 Kernthesen

Die Öffnung des Steuerbereichs für eine verstärkte und gezielte Nutzung von KI-Modellen beruht vor allem auf den aktuellen Entwicklungen generativer Sprachmodelle (z. B. Generative Pre-Trained Transformers (GPT)) und sogenannter Machine Learning Operations (MLOps) Plattformen. Diese stehen gerade deshalb im Fokus der aktuellen Studie, da Transformer-basierte Sprachmodelle [2] für weitreichende Verbesserungen in der maschinellen Text- und Sprachverarbeitung gesorgt haben. Mithilfe von MLOps ist auch der Steuerbereich mittlerweile in der Lage, den Lebenszyklus von KI-Modellen ganzheitlich zu optimieren und zu automatisieren. Damit können Lösungen entwickelt werden, die nicht nur KI-Experten und Entwickler, sondern zunehmend auch Fachexperten nutzen können.

Im Gegensatz zu Natural Language Processing (NLP) Methoden wie Recurrent Neural Networks (RNN) oder Long Short-Term Memory (LSTM) können Transformer-Modelle mittlerweile Kontexte in Textsequenzen über eine deutlich weitere Distanz aufrechterhalten. Auch eine parallele Datenverarbeitung ist möglich, wodurch das Trainieren und Anwenden deutlich effizienter abläuft. Ausgehend vom ursprünglichen Transformer-Modell wurden diverse neue Konzepte abgeleitet, wie z. B. BERT [3], welches ebenfalls ein zentraler Meilenstein im Bereich NLP darstellt. Aber auch allgegenwärtige Modelle wie GPT [4] basieren auf der Transformer-Architektur.

Eine weitere Entwicklung ist die Verfügbarkeit vortrainierter Modelle in Verbindung mit Konzepten wie dem Transfer Learning: Dabei werden Sprachmodelle zunächst auf einer großen allgemeinen Datenbasis vortrainiert, damit diese ein generelles Sprachverständnis erlernen. Anschließend können diese für bestimmte Aufgaben feinjustiert werden. Durch dieses Vorgehen wurde der Zugang und die Nutzung großer Sprachmodelle ermöglicht und vereinfacht. Dadurch können Unternehmen ein Sprachmodell auf ihre eigenen Anforderungen hin anpassen und sich so teils millionenschwere und aufwendige Trainingsprozeduren ersparen [5].

In den vergangenen Jahren lässt sich zudem ein starker Trend hin zu immer größeren Modellen (im Sinne ihrer trainierbaren Parameter) feststellen [6]. Heutzutage sind Modelle mit einer Parameteranzahl im dreistelligen Milliardenbereich keine Ausnahme mehr. In diesem Zusammenhang spricht man auch von Large Language Models (LLMs), die in Kapitel 2.3 näher vorgestellt werden. Auch wenn dies ein wesentlicher Grund für die stetig zunehmenden Trainingskosten ist, führt es auch dazu, dass die KI-Modelle eine immer bessere Performanz erzielen können. Zu welchen Leistungen solche Modelle im Steuerkontext in der Lage sind, zeigte beispielsweise eine Kurzstudie von Taxy.io, in der eine KI erstmalig die Steuerfachangestellten-Prüfung mit knapp 54% bestanden hätte [7].

So viel steht jedenfalls fest: Größere KI-Modelle sind heutzutage immer besser in der Lage, komplexere Zusammenhänge der natürlichen Sprache zu erfassen und zu verstehen. So lassen sich Texte mit zunehmend sprachlicher Qualität generieren und bspw. für kreative Schreibaufgaben nutzen. Trotz des durch ChatGPT Ende 2022 ausgelösten „Hypes“ zeigt der AI Index Report der Stanford University aus dem Jahr 2023 jedoch, dass die befragten Unternehmen generative Ansätze wie GAN (8%), NL Generation (12%) oder Transformer (11%) bisher nur begrenzt in ihre Unternehmenstätigkeiten integriert haben [6]. Aufgrund deren Neuartigkeit ist der Praxiseinsatz somit noch weitestgehend unerschlossen.

Auch im Kontext des Steuerbereichs befinden sich LLM-Anwendungen erst in den Anfangsphasen. Beispielsweise entwickelten Taxy.io und der Otto-Schmidt-Verlag erst kürzlich eine neue generative Steuer-KI, die es ermöglicht, Probleme hinsichtlich Halluzinationen, Bias und Fairness sowie Datenschutz entgegenzuwirken [8]. Eine richtungsweisende Entwicklung, die aber noch nicht breitflächig in der Praxis angekommen ist. Neben den Durchbrüchen generativer KI-Methoden zeigt sich aber auch, dass KI-Methoden nach wie vor für den Steuerbereich von hoher Relevanz sind, die schon seit längerer Zeit erforscht werden. Darunter fallen insbesondere Methoden zur Anomalieerkennung (vgl. Kapitel 2.6) und Klassifikation (vgl. Kapitel 2.7). Dies ist damit begründet, dass sich viele Anwendungsfelder im Steuerbereich auf diese Methoden übertragen lassen.

1.4 Zielsetzung

Ziel der vorliegenden Studie ist es, im Vergleich zur Vorgängerstudie aus dem Jahr 2017 die aktuellen Potenziale des Einsatzes von Künstlicher Intelligenz zur Unterstützung der Arbeit im Steuerbereich zu untersuchen. Dies geschieht vor dem Hintergrund der aktuellen Entwicklungen des Steuermarktes sowie der technologischen Durchbrüche im Bereich KI. Hierfür wurden in enger Zusammenarbeit zwischen Industrie, Beratung und Wissenschaft entsprechende Schlüsseltechnologien identifiziert, deren Anwendung bereits etabliert und/oder perspektivisch am vielversprechendsten für die Unterstützung von Aufgaben in der Steuerabteilung sind (vgl. Kapitel 2).

Durch die enge Zusammenarbeit von Industrie, Beratung und Wissenschaft kann zum einen sichergestellt werden, dass die relevanten Anwendungsfelder und Problemstellungen im Steuerbereich erhoben werden können. Und zwar basierend auf unmittelbarer Praxiserfahrung (vgl. Kapitel 3). Zum anderen lässt sich aber auch die prototypische Umsetzung mit Hilfe konkreter KI-Methoden wissenschaftlich fundiert realisieren und bewerten. Weiterhin sollen die Schlüsseltechnologien mit Blick auf ihre technischen und wirtschaftlichen Auswirkungen untersucht und bewertet werden.

Insgesamt lassen sich folgende Unterziele festhalten, die fester Bestandteil dieser Studie sind:

- › Schaffung eines Überblicks sowie die Einordnung und allgemeine Einführung von KI-Technologien, die für den Steuerbereich relevant sind und in dieser Studie fokussiert werden
- › Schaffung eines Überblicks über existierende KI-Anwendungen und KI-Trends im Steuerbereich
- › Ableitung und prototypische Umsetzung KI-basierter Anwendungsszenarien innerhalb der Steuerdomäne, um nicht nur die technischen und wirtschaftlichen Auswirkungen zu untersuchen und zu bewerten, sondern auch Handlungsempfehlungen für die Umsetzung entsprechender KI-Technologien zu geben

1.5 Methode

Um die Zielsetzung der Studie zu erreichen, wurden zunächst verschiedene Technologiebereiche als thematische Grundlage erfasst. Hierzu zählen all jene, die für den Steuerbereich von hoher Relevanz sind und im Rahmen der Studie näher betrachtet werden sollen. Die Auswahl erfolgte basierend auf einer Literaturanalyse und dem Austausch mit Fachexperten. Als zentrale Technologiebereiche wurden dabei große Sprachmodelle (LLM), Intelligente Roboterassistierte Prozessautomatisierung (iRPA) und Erklärbare KI (XAI) erfasst. Darüber hinaus werden allgemeinere Themen wie Maschinelles Lernen (ML), Klassifikationsalgorithmen, Anomalieerkennung sowie die Verwendung von synthetischen Daten betrachtet.

Die Relevanz dieser Bereiche ergibt sich unter anderem aus den bereits zuvor geschilderten Darstellungen und Entwicklungen. Dies bedeutet auch, dass die hier betrachteten Konzepte einer subjektiven, jedoch inhaltlich fundierten Vorauswahl unterliegen. Der Anspruch ist somit keine vollumfängliche Betrachtung sämtlicher KI-Teilbereiche und deren (denkbaren) Anwendungsmöglichkeiten für den Steuerbereich. Vielmehr handelt es sich um eine intensive Auseinandersetzung mit den oben geschilderten Technologien und welche Bedeutungen diese für den Steuerbereich im Sinne von möglichen Anwendungsszenarien haben. Dies geschieht insbesondere im Hintergrund der dargestellten technischen Entwicklungen seit der Vorgängerstudie.

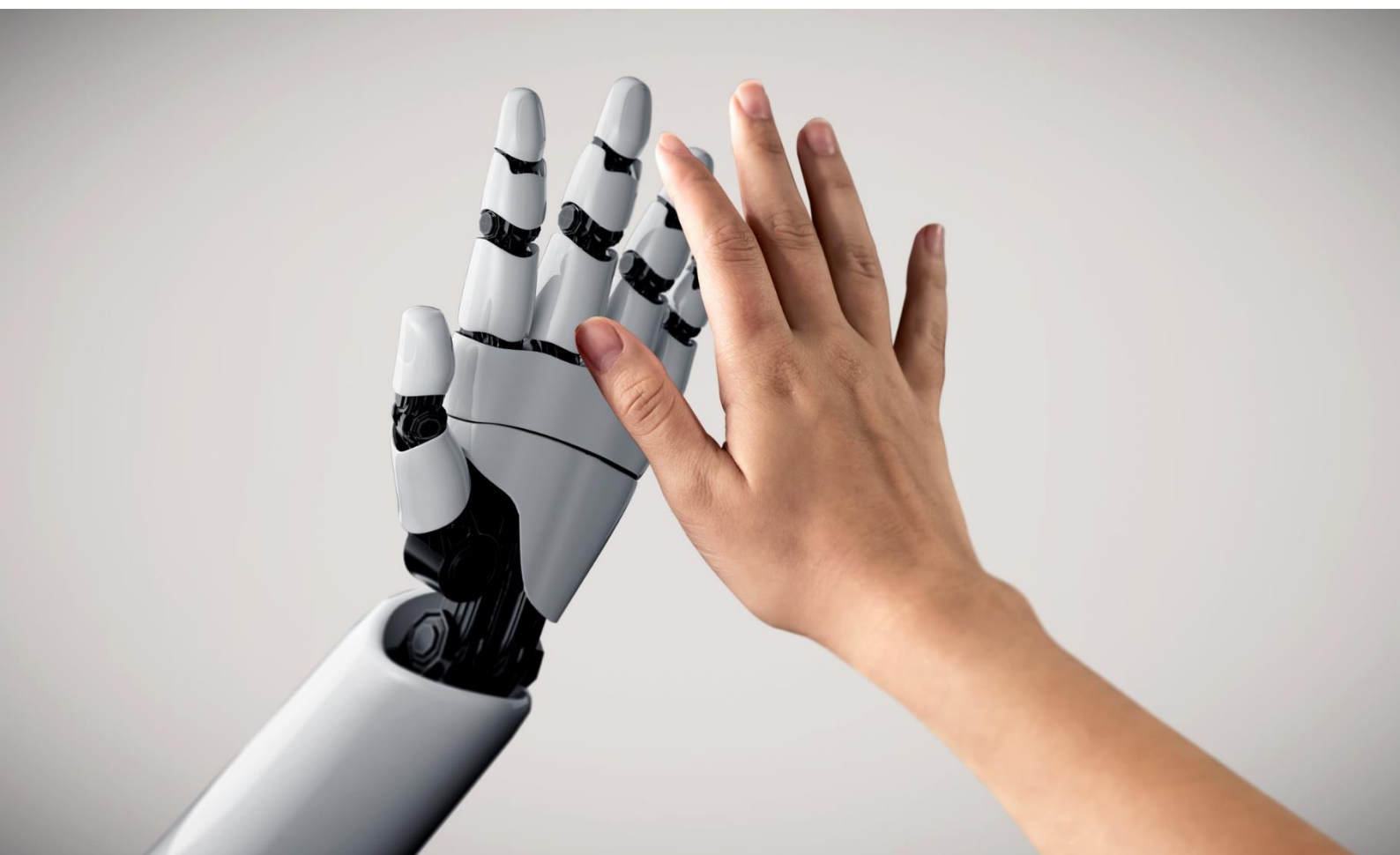
Um die betrachteten Technologiebereiche inhaltlich zu untermauern und für die Studie aufzubereiten, wurden Literaturrecherchen anhand wissenschaftlicher Arbeiten wie bspw. Konferenzbeiträge sowie Abhandlungen aus Fachzeitschriften durchgeführt. Ebenfalls wurden Publikationen aus Steuerberaterzeitschriften gesichtet, um den Stand von KI-Anwendungen im Steuerbereich aus einer praxisorientierten Sichtweise zu erfassen. Bei Letzterem waren weniger technologische Details und wissenschaftliche KI-Prototypen relevant, sondern die Erhebung von Meinungen und Erfahrungsberichten über Potenziale von KI-Anwendungen in der Steuerdomäne.

Darüber hinaus führte die WTS interne Experteninterviews mit Partnern und Fachexperten einzelner Steuerbereiche. Ziel dessen war es, zu erfassen, wie insbesondere die KI die Steuerabteilungen der Unternehmen in den letzten Jahren beeinflusst hat. Ein besonderer Fokus wurde in diesem Zusammenhang auf die Einsatzpotenziale und Herausforderungen gelegt, mit denen sich Steuerabteilungen in den kommenden Jahren von KI konfrontiert sehen. Auch Veränderungen der Arbeitswelt im Steuerbereich bedingt durch KI wurden diskutiert. Bei den Experteninterviews galt es, insbesondere Einblicke und Sichtweisen aus dem Praxisalltag zu erfassen.

Die so gewonnenen Impressionen konnten sowohl als Grundlage für die Bedeutung von KI im Steuerbereich in der Vergangenheit und Zukunft als auch für aktuell relevante Anwendungsfelder und Herausforderungen herangezogen werden. Ebenfalls wurden prototypische Implementierungen der als relevantesten eingeschätzten Anwendungsszenarien realisiert. Diese beziehen sich im Wesentlichen auf die zuvor definierten Technologiebereiche. Sie umfassen aber auch Anwendungsfälle aus Projekten von WTS, DFKI und weiteren externen Praxispartnern. Untersucht wurde dabei, wie und in welcher Form KI-Methoden für die angedachten Anwendungsfälle eingesetzt werden können. Entsprechende Lösungsarchitekturen wurden darauf basierend konzipiert und für einen Realwelteinsatz in prototypischen Demonstratoren eingebettet.

1.6 Aufbau und Struktur

Die vorliegende Studie ist wie folgt gegliedert: Aufbauend auf der allgemeinen Vorstellung der mit der Studie verfolgten Ziele und Motivation in Kapitel 1, dient Kapitel 2 der Darstellung thematischer Grundlagen der einzelnen Technologiebereiche, die für diese Studie relevant sind. Es wird sowohl eine allgemeine Einführung in den Themenkomplex KI und ML als auch tiefergehende Details der betrachteten Technologien geboten. Dabei wird auch ein Bezug zur steuerlichen Relevanz hergestellt. Im darauffolgenden Kapitel 3 wird ein Überblick über aktuelle Anwendungen von KI im Steuerbereich gegeben und darauf basierend ein aktueller Stand abgeleitet. Dies geschieht insbesondere durch eine Erhebung des Stimmungsbilds durch einen intensiven Austausch von steuerlichen Fachexperten. In Kapitel 4 werden Anwendungsszenarien dargestellt, die auf Basis der durchgeführten Mandanten-Workshops und Experteninterviews abgeleitet wurden. Dabei erfolgt eine detaillierte Beschreibung der allgemeinen Anwendungsfälle und der zugrundeliegenden Problemstellung, konkrete KI-Methoden und die erarbeiteten Lösungsarchitekturen sowie eine Darstellung der Ergebnisse. Kapitel 5 schließt die Studie mit einer Schlussbetrachtung ab und fasst die zentralen Ergebnisse noch einmal übergreifend zusammen.



2 Aktuelle Entwicklungen der Künstlichen Intelligenz

2.1 Einführung

Die KI kommt seit Jahrzehnten nicht nur im privaten, sondern auch im wirtschaftlichen und öffentlichen Bereich zum Einsatz. Spätestens aber mit der Veröffentlichung von ChatGPT ist das Thema auch in der breiten Öffentlichkeit angekommen. Wie kaum ein anderer Trend wecken vor allem generative KI-Ansätze neue Hoffnungen. Denn dadurch ergeben sich bisher ungeahnte Möglichkeiten zur Datenverarbeitung und zur automatisierten Generierung neuer Inhalte. Zugleich schüren die neuen Hilfsmittel aber auch berechtigte Ängste. Unabhängig davon hat sich die KI in den letzten Jahren enorm weiterentwickelt und kann auf eine lange Historie zurückblicken.

Entwicklung von KI

Der Begriff der KI geht auf John McCarthy zurück. 1955 definierte er das Wissensgebiet der KI als einen Prozess, in dem es darum geht, „Maschinen zu entwickeln, die sich verhalten, als verfügten sie über menschliche Intelligenz“ [9]. Dieses Zitat stammt aus einer Zeit, als man glaubte, mithilfe von Tests, Intelligenz und den sogenannten Intelligenzquotienten genau ermitteln zu können. Dazu hatte Alan Turing im Jahr 1950 in einem berühmten Artikel ein Gedankenexperiment vorgestellt. Mit Hilfe des sogenannten „Turing-Test“ versuchte der britische Mathematiker, die KI messbar zu machen. Ziel des Tests war es herauszufinden, wer von zwei Gesprächspartnern ein Mensch und wer ein Chatbot ist.

Mit Blick auf ihre Anfänge in der Mitte des 20. Jahrhunderts hat die KI mehrere zentrale Entwicklungsphasen durchlaufen. Die wichtigsten davon sind im ungefähren Zeitrahmen in Abbildung 1 dargestellt (vgl. [2], [9]–[13]):

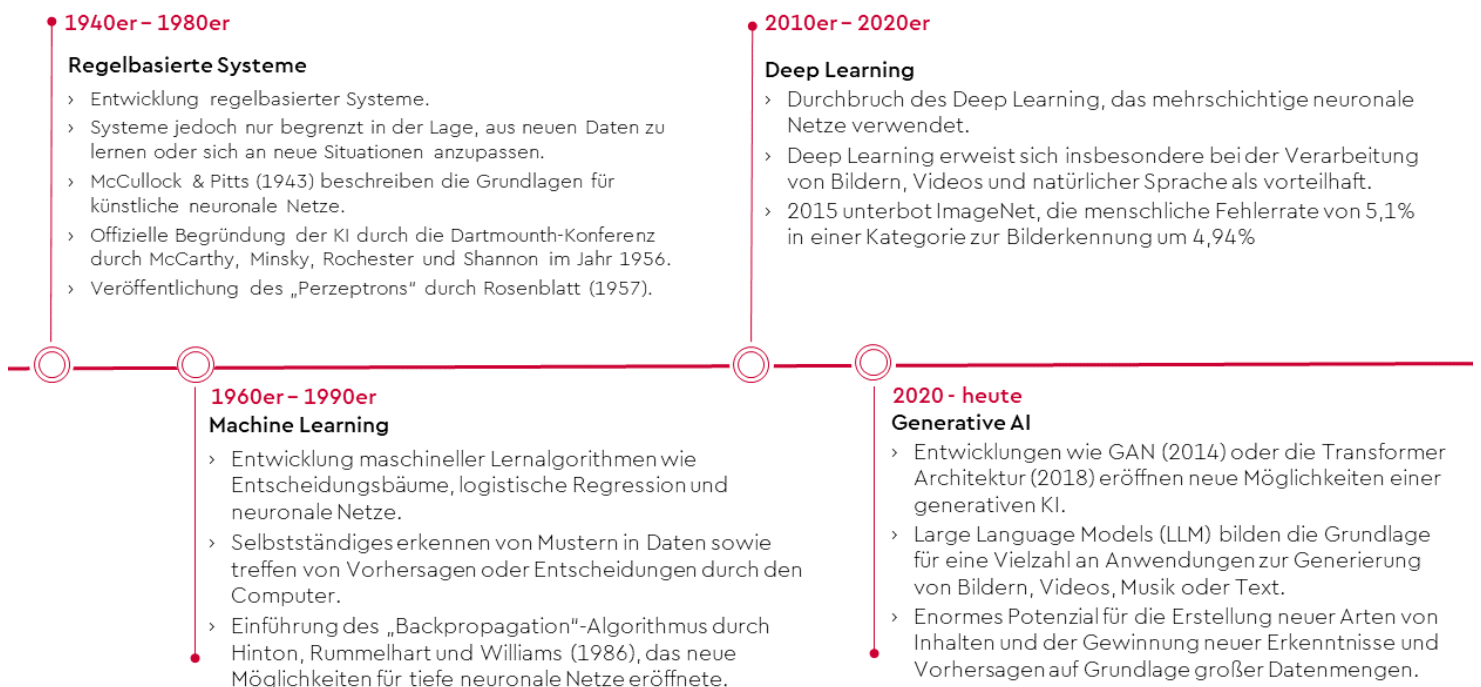


Abbildung 1: Entwicklung der Künstlichen Intelligenz

Bedeutung von KI für die Steuerfunktion

Die Steuerdomäne ist in vielerlei Hinsicht kompliziert, baut zugleich aber auch sehr auf systematischen und logischen Prozessen auf. Ein Großteil der Arbeit besteht in der Einhaltung von Vorschriften und in der Datenbereinigung bzw. -verarbeitung. Allein deshalb eignet sich eine Vielzahl von Tätigkeiten ideal für den Einsatz von KI-Technologien. Hierzu zählen bspw. natürliche Sprachverarbeitung, intelligente Prozessautomatisierung, musterbasierte Datenanalyse und die Sicherstellung von Transparenz. Die Vorteile liegen klar auf der Hand: Einerseits lassen sich alte Dienstleistungen im Steuerwesen auf neue Art und Weise erbringen. Zum anderen eröffnen neue KI-Technologien aber auch völlig neue Möglichkeiten. So entstehen Dienstleistungen, an die bisher noch gar nicht gedacht wurde oder die technisch noch nicht möglich waren.

Vor diesem Hintergrund beleuchtet die vorliegende Studie das Anwendungspotenzial ausgewählter KI-Technologien für den Steuerbereich. Jede einzelne dieser Technologien bzw. jeder Ansatz wird kurz vorgestellt und deren Bedeutung bzw. deren Nutzen für den Steuerbereich näher beleuchtet – unter Berücksichtigung der jeweils damit verbundenen Grenzen und Einschränkungen.

2.2 Maschinelles Lernen

Kurzvorstellung

Maschinelles Lernen (engl. Machine Learning (ML)) befasst sich in erster Linie mit der Entwicklung selbstlernender Algorithmen, die aus großen Datenmengen neue Erkenntnisse extrahieren können. Immer mit dem Ziel, eventuell vorhandene Muster zu erkennen oder Vorhersagen über zukünftig eintretende Ereignisse zu treffen. Die Vorstellung, wie die Algorithmen dabei lernen, ist ähnlich zu den Fähigkeiten eines Menschen. Die Maschine startet mit einem rudimentären Weltverständnis analog zur Entwicklungstheorie „tabula rasa“ des britischen Philosophen John Locke. Anhand von eintretenden Ereignissen werden neue Erkenntnisse gesammelt und Rückschlüsse gebildet, um diese im Anschluss neu miteinander zu verknüpfen. In der Folge wird das Verständnis für ein ausgewähltes Anwendungsgebiet kontinuierlich weiterentwickelt. Hierzu zählen bspw. die automatisierte Identifikation von SPAM-Mails, die Untersuchung von Wörtern, Sätzen oder ganzen Texten im Zuge sogenannter Sentiment-Analysen (z. B. Twitter-Analysen), die Betrugserkennung in Transaktionsdaten (Fraud Detection), Empfehlungssysteme, Chatbots und Sprachassistenten wie Siri, Alexa und Google Now [1].

Allen Ansätzen und Methoden des ML ist gemein, dass Algorithmen nach Mustern in großen Datenmengen suchen. Auf deren Basis werden dann neue Informationen abgeleitet, Zusammenhänge innerhalb dieser erkannt und schließlich eigenständig Entscheidungen getroffen. Die Verfahren lassen sich dabei nach Art und Weise der Zielsetzung und Funktionsweise in verschiedene Rubriken kategorisieren. Grundsätzlich unterscheidet man zwischen drei Lernverfahren [1]:

Überwachtes Lernen (engl. Supervised Learning):

Dem Algorithmus werden Beispieldaten mit bekannten Eingaben und den dazugehörigen Ausgaben zur Verfügung gestellt, um eine Funktion zu approximieren. Der Algorithmus entwickelt also eine möglichst simple Rechenformel, um mit deren Hilfe die verschiedenen Eingaben den richtigen Ausgaben zuzuordnen zu können.

Unüberwachtes Lernen (engl. Unsupervised Learning):

In diesem Fall werden dem Algorithmus nur die Eingabedaten ohne die dazugehörigen Ausgabewerte, sog. *Labels*, gegeben. Anschließend sucht der Algorithmus in den vorhandenen Datenmengen dann selbstständig nach Strukturen und Mustern, um neue Erkenntnisse zu gewinnen.

Bestärkendes Lernen (engl. Reinforcement Learning):

Anders als bei den beiden zuvor genannten Lernverfahren interagiert der Algorithmus (Agent) hierbei mit einer Umgebung (Environment) und erhält Rückmeldungen bzw. Feedbacks (Belohnungen oder Bestrafungen) basierend auf seinen Aktionen. Der Algorithmus lernt kontinuierlich per Versuch und Irrtum (Trial and Error) die besten Aktionen auszuwählen, um ein bestimmtes Ziel zu erreichen. Im Gegensatz zum Überwachten Lernen handelt es sich beim Feedback-Verfahren aber nicht um korrekte Klassenbezeichnungen, sondern um ein Maß, wie gut die Aktion im Ausführungskontext war. Einen Überblick über typische Verfahren ist in Abbildung 2 dargestellt.

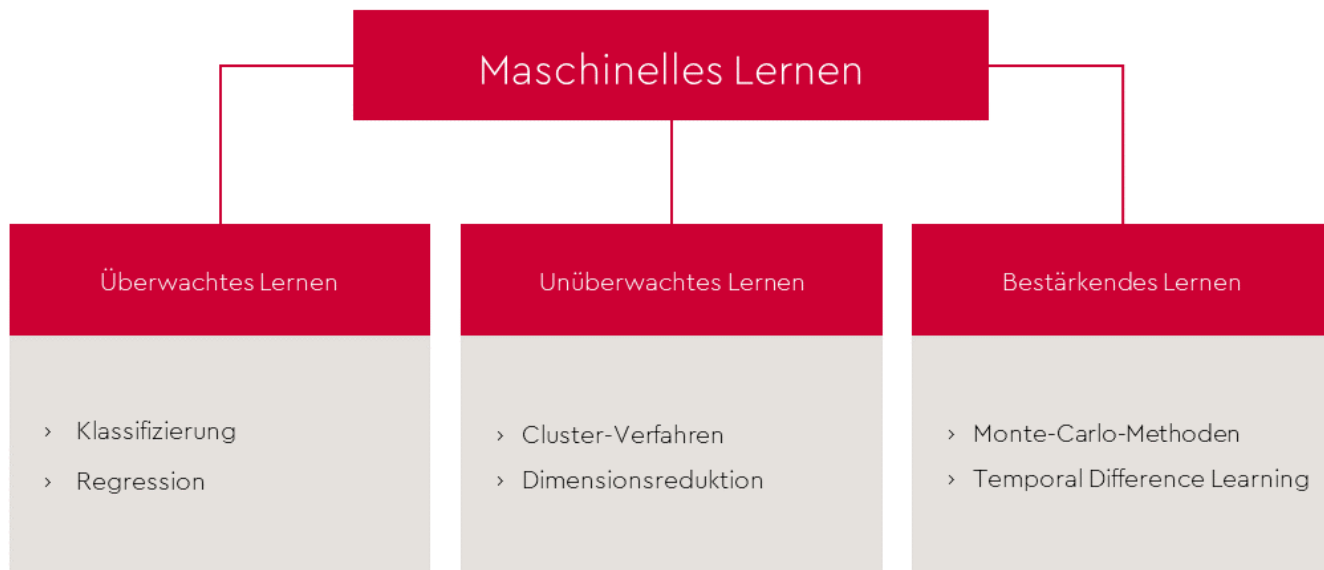


Abbildung 2: Übersicht Maschinelles Lernen [1]

Bedeutung und Nutzen für den Steuerbereich

ML ist in dreierlei Hinsicht relevant für den Steuerbereich. Zum einen müssen große Mengen an Finanz- und Geschäftsdaten verarbeitet und analysiert werden. Des Weiteren muss die Steuerabteilung mit verschiedenen Steuerarten, Gesetzen und Compliance-Richtlinien umgehen können. Aber auch mit wiederkehrenden Aufgaben wie der Erstellung und Überprüfung von Steuererklärungen, der Identifizierung von Abweichungen und der Einhaltung von Fristen ist die Steuerabteilung konfrontiert.

Vor dem Hintergrund eines effektiven Einsatzes von Personalressourcen kann ML die Steuerfunktion bei der Identifizierung potenzieller Betrugsfälle oder Fehler im Rahmen der Steuerprüfung unterstützen. Indem es Muster und Anomalien in großen Datenmengen erkennt, lassen sich verdächtige Transaktionsmuster in Echtzeit überwachen und im Fall von ungewöhnlichen Aktivitäten entsprechende Alarme auslösen. Aber auch die Genauigkeit und Effizienz der Erstellung von Steuererklärungen kann durch ML erhöht werden. Hierfür kann das Modell sämtliche Steuerdaten durchsuchen und automatisch die entsprechenden Steuerabzüge und -anpassungen berechnen. Eine Eigenschaft, die letztendlich zu einer Verbesserung der steuerlichen Compliance führt. Auch wiederkehrende und zeitaufwendige Aufgaben, wie z. B. die Überprüfung von Steuererklärungen oder die Extraktion von relevanten Informationen aus Dokumenten, lassen sich durch ML automatisieren. In Kombination mit LLMs können sogar neue Services entwickelt werden, um bspw. konkrete Fragen zu einem Dokument automatisch zu beantworten.

Grenzen für die Nutzung im Steuerbereich

Neben den genannten Vorteilen birgt der Einsatz von ML aber auch gewisse Limitationen. So basieren die trainierten Modelle zumeist auf bereits vorhandenen historischen Daten. Sind diese nicht aktuell, vollständig oder sogar fehlerhaft, kann das Modell Schwierigkeiten haben, komplexe steuerliche Ausnahmeregelungen oder rechtliche Nuancen zu verstehen. Ganz im Gegenteil zu menschlichen Experten, die diese Punkte allein aufgrund ihrer Erfahrung berücksichtigen würden. Der Einsatz von ML erfordert zudem den Zugriff auf große Mengen sensibler, steuerlicher und finanzieller Daten. Ebenfalls können Probleme auftreten, sollte sich die Qualität, Relevanz und Verfügbarkeit von Trainingsdaten bspw. zu bestimmten steuerlichen Szenarien nicht sicherstellen lassen. In diesem Fall ist die Fähigkeit des Modells beeinträchtigt, genaue Vorhersagen oder Entscheidungen in Bezug auf diese Szenarien zu treffen.

Die Verwendung eines vortrainierten Modells und dessen Anwendung auf eine neue Aufgabe kann durchaus als Lösungsalternative angesehen werden. Dennoch bedarf es insbesondere im Steuerbereich weiterer Forschung und Entwicklung, inwiefern Transfer Learning für bestimmte Aufgaben überhaupt eingesetzt werden kann. Zudem sind steuerliche Entscheidungen oft mit erheblichen finanziellen Auswirkungen verbunden. Aus diesem Grund müssen die dem trainierten Modell zugrundeliegenden Faktoren und Logiken für Steuerexperten nachvollziehbar sein, damit diese die Richtigkeit der Ergebnisse überprüfen und ggf. korrigieren können. Je nach Komplexität des Modells könnte es bspw. im Kontext von Verrechnungspreisen schwierig sein, der jeweiligen Finanzbehörde zu erklären, warum ein bestimmter Preis oder (bestimmte) Methode durch das KI-Modell vorgeschlagen und in die Verrechnungspreisdokumentation mit aufgenommen wurde.

2.3 Große Sprachmodelle

Kurzvorstellung

Große Sprachmodelle (engl. Large Language Models (LLM)) sind fortschrittliche KI-Modelle, welche die Verarbeitung natürlicher Sprache revolutioniert haben. So sind Modelle wie z. B. Generative Pre-trained Transformer (GPT) in der Lage, menschenähnliche Texte zu verstehen und zu generieren. Hierfür werden sie gezielt auf großen Datenmengen trainiert. Darüber hinaus kommen Deep-Learning-Architekturen zum Einsatz, um komplexe Sprachmuster, semantische Beziehungen und Kontext zu erfassen.

Insbesondere die Transformer-Technologie ermöglicht es großen Sprachmodellen, weitreichende Abhängigkeiten und kontextuelle Beziehungen zwischen Wörtern in einer Textfolge zu erfassen. Dazu nutzen Transformer den sogenannten Self-Attention-Mechanismus. Dieser macht es ihnen möglich, jedem Wort in einem Satz eine unterschiedliche Gewichtung oder Bedeutung zuzuweisen – stets auf der Grundlage seiner Relevanz für andere Wörter [2].

Die eigentliche Besonderheit von großen Sprachmodellen liegt in ihrer Fähigkeit, kohärenten und kontextuell relevanten Text zu erzeugen. Ebenso zeichnen sie sich durch ein hohes Verständnis von Vokabular und Grammatik aus und zeigen sogar ein gewisses Maß an Kreativität. Durch ihre Generalität bieten sie große Potenziale für zahlreiche Anwendungen wie z. B. Chatbots, Text Mining oder digitale Assistenten (z. B. Alexa oder Siri). Beim sogenannten Natural Language Processing (NLP) handelt es sich um eine Sammlung von KI-Methoden, die es Computern ermöglichen, die natürliche Sprache des Menschen zu verstehen und entsprechend zu verarbeiten.

Bedeutung und Nutzen für den Steuerbereich

Große Sprachmodelle sind vor allem aufgrund ihrer Fähigkeiten zur Sprachverarbeitung von großer Bedeutung für die Steuerfunktion. Denn einerseits umfasst der Steuerbereich eine komplexe Terminologie mit einem besonderen juristischen Fachjargon. Andererseits existiert eine große Menge an Vorschriften und Richtlinien, die identifiziert und korrekt interpretiert werden müssen. Nur so lassen sich richtige Entscheidungen treffen. Im Folgenden werden drei Anwendungsfelder aus dem Steuerbereich skizziert, in denen Sprachmodelle zum Einsatz kommen können:

- (1) Steuereinhaltung:** Große Sprachmodelle können der Steuerfunktion helfen, komplexe Steuervorschriften und -richtlinien zu verstehen. Sie können genaue und aktuelle Informationen zu Steuergesetzen, Abzügen, Gutschriften und Compliance-Anforderungen bereitstellen und die Einhaltung von Steuervorschriften gewährleisten.
- (2) Steuerplanung und -optimierung:** Große Sprachmodelle können Finanzdaten analysieren und Einblicke in Steuerplanungsstrategien bieten. Durch die Verarbeitung von Informationen über Einkommen, Ausgaben, Investitionen und geltende Steuergesetze können sie zudem Empfehlungen zur Optimierung der steuerlichen Ergebnisse geben und potenzielle Abzüge oder Gutschriften ermitteln.
- (3) Analyse von Steuerdokumenten:** Große Sprachmodelle können bei der Analyse von Steuerdokumenten wie Jahresabschlüssen, Quittungen und Rechnungen helfen. Ebenso können sie relevante Informationen extrahieren, Dokumente klassifizieren und dabei helfen, den Prozess der Steuererstellung zu automatisieren, um Zeit zu sparen und Fehler zu reduzieren.

Die erfolgreiche Anwendung von großen Sprachmodellen im Steuerbereich erfordert ggf. ein Prompt-basiertes Tuning oder ein zusätzliches Fine-Tuning des Modells. Bei Ersterem wird das Modell nicht verändert, sondern sein zuvor trainiertes Wissen genutzt, indem spezifische Aufforderungen zur Steuerung des Antwortverhaltens verwendet werden. Im Gegensatz dazu befasst sich das Fine-Tuning mit einem erneuten Training des LLM auf einem spezifischen Datensatz, wodurch die Modellparameter geändert werden, um es auf den zugrundeliegenden Datensatz und die damit verbundene Aufgabe zu spezialisieren.

Grenzen für die Nutzung im Steuerbereich

Große Sprachmodelle bieten ein beträchtliches Potenzial für den Steuerbereich. Dennoch ist es wichtig, auch deren Grenzen zu beachten, da auch gelegentlich falsche oder verzerrte Informationen erzeugt werden können. Unterscheiden sich im Laufe der Zeit zudem die Eingabe- von den Trainingsdaten, kann es zudem zu einem sog. *model drift* kommen. Ein Phänomen im Umfeld des maschinellen Lernens, das zu einer Verschlechterung der Genauigkeit des KI-Modells führen kann.

Dementsprechend wichtig ist die Sicherstellung einer hohen Datenqualität im Sinne sauberer, gut beschrifteter und repräsentativer Daten. Auch die Robustheit des Sprachmodells muss sichergestellt sein. Nicht zu vergessen die Komplexität, aufgrund derer sowohl das Training als auch die Ausführung erhebliche Rechenressourcen benötigen. In diesem Zusammenhang kann allein der fehlende Zugang zu leistungstarker Hardware zur echten Herausforderung werden.

Ferner gibt es bei der Verwendung einiger generativen KI-Modelle berechtigte Bedenken, was das Eigentum an den Ergebnissen und dem geistigen Eigentum betrifft. Denn diese Art künstlicher Intelligenz ist in der Lage, aus bereits bestehenden Datensätzen neue Inhalte und Ideen zu erzeugen. Dementsprechend kritisch ist die Beurteilung hinsichtlich der Zuverlässigkeit der Trainingsdaten und des Urheberrechts bzw. der Quellenzuordnung. Die für das Training öffentlicher Modelle verwendeten Daten sind zwar umfangreich, basieren aber möglicherweise auf unvollständigem oder veraltetem Wissen. Dadurch lassen sich Fakten nicht immer exakt überprüfen bzw. müssen bei der Verwendung ggf. noch einmal kontrolliert oder recherchiert werden.

Des Weiteren stellt der Datenschutz einen weiteren wichtigen Aspekt dar. Diesen gilt es nicht nur im Rahmen des Modelltrainings, sondern auch bei der aus einer bestehenden Datenbasis gezogenen Schlussfolgerung (Inferenz) zu berücksichtigen. Werden beispielsweise große Sprachmodelle von externen Dienstleistern bezogen, müssen möglicherweise vertrauliche Daten (z. B. Kundendaten) mit diesem geteilt werden. Wie die Vertraulichkeit der Daten innerhalb der Modelle geschützt werden kann, ist allerdings eine noch offene Forschungsfrage. Aufgrund der probabilistischen Struktur muss man sich zudem bewusst sein, dass die Qualität der Vorhersagen je nach Eingabe variieren kann. Mögliche Gründe hierfür können u. a. die verwendeten Trainingsdaten, die Eingabestruktur oder unterschiedliche Parametereinstellungen bei der Nutzung der Modelle sein.

2.4 Intelligente Robotergestützte Prozessautomatisierung

Kurzvorstellung

Der Begriff der robotergestützten Prozessautomatisierung (engl. Robotic Process Automation (RPA)) vereint Robotik in Form von Software-Agenten (Bots) mit der Automatisierung von Prozessen. Dazu nutzen Software-Agenten bestehende Software-Anwendungen [14] und können damit wie Menschen interagieren. Indem sie Programme wie bspw. ERP- und CRM-Systeme aufrufen, innerhalb der Programme navigieren, Informationen extrahieren, Eingaben tätigen, um Informationen zu übertragen, Knöpfe drücken und zwischen verschiedenen Programmen hin und her wechseln, sind sie in der Lage, einen zuvor definierten Arbeitsablauf aus der Anwendersicht auszuführen.

Die von einem sogenannten Bot ausgeführten Aufgaben erfolgen zumeist regelbasiert, gut strukturiert und wiederkehrend bzw. repetitiv [15]. Durch die Automatisierung sollen Arbeitsabläufe beschleunigt, skaliert und letztendlich auch rationalisiert werden. Das Ergebnis sind freiwerdenden Mitarbeiter, die sich auf weitaus wichtigere Aufgaben mit höherer Wertschöpfung konzentrieren können[16]. Die kontinuierliche Integration von KI-Technologien wird auch als intelligente robotergestützte Prozessautomatisierung (iRPA) bezeichnet (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**).

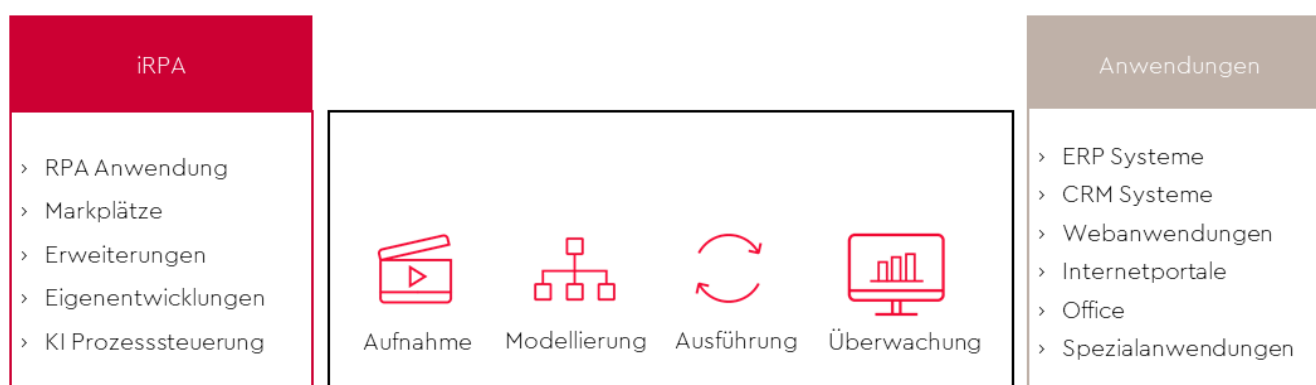


Abbildung 3: Überblick über iRPA

iRPA kann dabei zwei Bedeutungen haben:

- (1) **Ausführende Bots:** Selbstständiges Erlernen menschlicher Interaktionen mit Anwendungssystemen sowie neuer Verarbeitungsmethoden durch die Software.
- (2) **Einsatz von KI-Methoden:** Unterstützung dedizierter Prozessschritte durch KI-technologien wie bspw. die Erkennung und Verarbeitung von Bilddaten oder gescannten Texten durch Optical Character Recognition (OCR) sowie die automatisierte Ableitung von Empfehlungen aus unterschiedlichen Informationsquellen.

Bedeutung und Nutzen für den Steuerbereich

Generell eignen sich sowohl die RPA als auch iRPA für sich wiederholende Aktivitäten, bei denen automatisch vordefinierte Prozesspfade durchlaufen werden. Mit iRPA werden zudem komplexe Berechnungen und die Verarbeitung von unstrukturierten Daten unterstützt. Im steuerlichen Kontext betrifft das oftmals folgende Aspekte [17]:

- (1) Datenbeschaffung:** Datenquellen wie bspw. Papierdokumente oder gescannte Dokumente, E-Mails, Sprachkommunikation, die Erfassung über externe Netzwerke und digitale Schnittstellen können mittels Deep Learning, Bildverarbeitung, Texterkennung (OCR) und dem Internet der Dinge maschinell erschlossen werden. Hierfür müssen diese nicht einmal händisch in ein digitales Format übertragen werden. Auch das Auslesen von Webseiten oder Tabellen für den automatischen Übertrag in ein elektronisches Datensystem zur weiteren Verarbeitung eignen sich dafür [18].
- (2) Datenverarbeitung:** Zu den konventionellen Verarbeitungsmethoden zählen die Prüfung von Informationen, die Überprüfung von Regeln, die Eingabe von Anfragen, das Einfügen von Kopien sowie das Wechseln von Anwendungen oder Seiten. Mit der Integration von KI-Anwendungen können RPA-Bots nicht nur zur Automatisierung durch Klickfolgen eingesetzt werden. Auch die Verarbeitung natürlicher Sprache, die Klassifikation von Daten oder die Erkennung von Datenmustern lassen sich in die Abläufe integrieren. Im steuerlichen Bereich sind das z. B. die Bereinigung und die Vereinheitlichung von gesammelten Daten durch die Entfernung von Duplikaten und Inkonsistenzen. Ebenso möglich ist auch der Abgleich von Finanzdaten aus verschiedenen Quellen wie z. B. Kontoauszügen, Rechnungen und Spesenabrechnungen.
- (3) Entscheidungsfindung:** Zu den traditionellen Methoden zählen Regelsysteme, Datenanalyse, Sammlung von Erfahrungen und mehrdimensionale Datenanalyse. Mit Einsatz von KI-Anwendungen kommen in vielen Aspekten insbesondere Empfehlungssysteme, prädiktive und präskriptive Analysen, Stream-Processing und Big-Data-Analyse mit hinzu. Im steuerlichen Bereich sind das z. B. die Identifikation und der Umgang mit Diskrepanzen, Fehlern, Unstimmigkeiten oder Abweichungen in Daten und Transaktionen. So lassen sich bspw. Einnahmen- und Ausgabenaufzeichnungen abgleichen, Steuerdokumente gegenprüfen und Jahresabschlüsse abstimmen.
- (4) Überwachung:** Zu den konventionellen Methoden zur Überwachung von Prozessen gehören Daten- und Ergebnisanalyse, Regelüberprüfung, mehrdimensionale Analysen sowie Optimierung und Verbesserung. Bei der Überwachung von Prozessen durch KI-Systeme kommen eher Risikovorhersagen, Deep Learning und maschinelles Lernen, Videoanalyse und die Verarbeitung natürlicher Sprache zum Einsatz. Denn mit all diesen Verfahren lässt sich die Ausführung von Prozessen und Aktivitäten hinsichtlich Qualität und Performanz bequem überwachen. Im steuerlichen Bereich sind das z. B. die permanente Überwachung von Finanzdaten hinsichtlich etwaiger Muster, Trends und Ausreißer sowie die Kategorisierung von Transaktionen auf der Grundlage vordefinierter Regeln. Diese Analyse kann dabei helfen, potenzielle Steuerprobleme oder Bereiche mit Optimierungsbedarf zu identifizieren. Auf diese Weise lassen sich Warnmeldungen oder Berichte über nicht konforme Transaktionen erstellen, damit die zuständigen Steuerberater umgehend Korrekturmaßnahmen ergreifen können.

Grenzen für die Nutzung im Steuerbereich

Es existiert noch immer kein einheitliches Verständnis darüber, was unter dem Begriff RPA zu verstehen ist – und was nicht. Insbesondere in der Literatur wird der Begriff deutlich weiter gefasst, als es in der Praxis der Fall ist. RPA wird eher als ein Sammelbegriff für eine Vielzahl von Methoden verstanden, die eine automatisierte Prozessausführung unter Verwendung bestehender Anwendungssysteme ermöglichen [19].

Die konkrete Implementierung von iRPA ist zudem mit einigen Herausforderungen verbunden. Doch viele davon resultieren unmittelbar aus der KI-Integration selbst [20]. Hierzu zählen bspw. der Umfang und die Qualität von Trainingsdaten sowie die (oft) mangelnde Erklärbarkeit von Vorhersagen/Entscheidungen der KI-Modelle (vgl. Kapitel 2.5). Die führenden Anbieter von RPA-Lösungen bieten weiterhin nur geringfügig kognitive Fähigkeiten im standardmäßigen Funktionsumfang. Stattdessen müssen entsprechende Komponenten meist über Drittanbieter durch den Endanwender integriert werden [21]. Dies kann sich bspw. dadurch äußern, dass die RPA-Anbieter eine Art Marktplatz anbieten, in welchem kognitive Komponenten bereitgestellt und direkt in die RPA-Skripte importiert werden können [22]. Andernfalls besteht auch die Möglichkeit, KI-Modelle unabhängig von der RPA-Lösung zu erstellen, die dann von einem RPA-Skript gezielt abgerufen und verwendet werden können.

2.5 Erklärbare Künstliche Intelligenz

Kurzvorstellung

Die erklärbare künstliche Intelligenz (engl. eXplainable AI (XAI)) bezeichnet einen Forschungsbereich, der sich mit Methoden und Ansätzen beschäftigt, mit deren Hilfe sich Entscheidungen und Vorhersagen von komplexeren KI-Modellen transparenter und nachvollziehbarer machen lassen. Man spricht in diesem Zusammenhang auch von der Erklärung von „Black-Box-Modellen“. Entsprechende Ansätze sind vor allem dann relevant, wenn ihnen KI-Modelle in Form von Neuronalen Netzen zugrunde liegen, deren Ausgaben keine nachvollziehbaren Rückschlüsse auf die Entscheidungen ermöglichen.

XAI kann aber auch dafür verwendet werden, ein erstelltes KI-Modell auf dessen Performanz hin zu untersuchen, d.h. ob die Modellausgaben auch einer manuellen Beurteilung entsprechen würden und dementsprechend nachvollziehbar sind. Erkenntnisse aus diesen Analysen könnten dann für potenziell notwendige Modellanpassungen herangezogen werden. Grundsätzlich wird im Zusammenhang mit XAI darin unterschieden, ob direkt erklärbare Modelle wie Entscheidungsbäume oder Black-Box-Modelle zum Einsatz kommen.

- (1) Entscheidungsbäume:** Dabei handelt es sich um ein perfektes Beispiel für direkt erklärbare Modelle. Denn deren Entscheidungen im Sinne von Klassifikationen, die sie vornehmen, gehen unmittelbar aus der Baumstruktur hervor. Somit wird klar, welche Attribute und deren Ausprägungen schrittweise zur finalen Klassifikation geführt haben.
- (2) Black-Box-Modelle:** Dabei handelt es sich um sog. „post-hoc Erklärungen“, da entsprechende Erklärungslösungen im Anschluss der Vorhersagen ihre Anwendung finden. Hierbei unterscheidet man wiederum, ob das Gesamtmodellverhalten erklärt werden soll oder nur einzelne Vorhersagen (im Sinne einzelner Datenpunkte). Für ersteres sind globale XAI-Methoden relevant, während letzteres durch sog. lokale Erklärungsmethoden realisiert wird. Zu den in der Forschung und in der Praxis am weitesten verbreiteten XAI-Methoden zählen LIME („Local Interpretable Model-agnostic Explanations“) und SHAP („SHapley Additive exPlanations“).

Die von XAI-Ansätzen produzierten Ergebnisse lassen sich problemlos anschaulich visualisieren. Für einzelne Datenattribute ist aber auch eine Darstellung möglich, die den quantitativen Einfluss der XAI-Ansätze auf die finale Vorhersage aufzeigt. Dabei kann bspw. visualisiert werden, ob niedrige oder hohe Attributausprägungen zu einem negativen bzw. positiven Einfluss auf die Vorhersage geführt haben.

Ein konkretes Beispiel für solche Erklärungen ist ein KI-Modell, das basierend auf diversen Attributen eines Grundstücks dessen Marktwert vorhersagen soll. In diesem Fall ist es naheliegend, dass z. B. die Grundstücksfläche einen positiven, sprich erhöhenden Einfluss auf den Marktwert hat. Dabei ist zu erwarten, dass mit zunehmender Fläche auch der Marktpreis entsprechend steigt. Visualisiert man ein solches Vorhersagemodell mit den geschilderten XAI-Methoden, können die Zusammenhänge bestätigt werden.

Mit Hilfe von XAI können aber auch wichtige Faktoren wie *Fairness* und *Bias* (Vorurteil) eines KI-Modells adressiert werden. Denn insbesondere falsche Untersuchungsmethoden wie bspw. Suggestivfragen führen mitunter zu verzerrten oder sogar falschen Ergebnissen.

Zur besseren Veranschaulichung ein praktisches Beispiel: Ein kritischer Anwendungsfall ist die Entwicklung eines Klassifikationsmodells, das die Kreditwürdigkeit einer Person anhand verschiedener personenbezogener Attribute einstufen soll. Denkbare Merkmale könnten neben der Höhe des Kredits bspw. das Geschlecht oder die Religionsangehörigkeit sein. Trainiert man ein solches Modell anhand von vorliegenden Daten, die einen Bias gegenüber bestimmten Personengruppen enthalten, würde sich dies unmittelbar auch auf das Modell auswirken. Eine entsprechende Erklärung des Modells mit Hilfe von XAI-Methoden könnte genau solche Problematiken aufdecken. Indem für die einzelnen Attribute aufgezeigt wird, welche Ausprägung tatsächlich zu welchen Entscheidungen des Modells führen.

Bedeutung und Nutzen für den Steuerbereich

Für den Steuerbereich ist XAI in vielerlei Hinsicht relevant: So nimmt die Anzahl an KI-Anwendungen im Steuerbereich stetig zu. Insbesondere komplexere KI-Methoden kommen dabei immer häufiger zum Einsatz, während „einfachere“ Verfahren des ML (wie bspw. Entscheidungsbäume, die eine transparente Erklärung für die Vorhersagen bieten) eine geringere Beliebtheit erfahren. Somit lässt sich ein Trend zu immer performanteren, jedoch zunehmend intransparenteren Methoden feststellen. In diesem Zusammenhang können XAI-Methoden dazu beitragen, für Unternehmenstätigkeiten verwendete intransparente KI-Modelle besser zu erklären und deren Entscheidungen in nachvollziehbarer Form darzulegen.

Gerade der Steuerbereich unterliegt diversen Compliance-Anforderungen und gesetzlichen Regelungen. Dies ist mit ein Grund, warum steuerrelevante Entscheidungen sowohl eine Kontrollierbarkeit nach innen als auch nach außen ermöglichen sollten. Außerdem regeln diverse Gesetze und Rechte die Behandlung des Individuums, die auch dessen Besteuerung betreffen. Hier sei beispielhaft die Europäische Menschenrechtskonvention mit Artikel 6 („Recht auf ein faires Verfahren“), Artikel 8 („Recht auf Achtung des Privat- und Familienlebens“) und Artikel 14 („Diskriminierungsverbot“) genannt[23]. Auch die DSGVO regelt bspw. in Artikel 22, wie und ob Personen basierend auf automatisierten Entscheidungen behandelt werden, wie es bspw. KI-Lösungen ermöglichen.

Grenzen für die Nutzung im Steuerbereich

Auch wenn sich die Erklärbarkeit von Black-Box-Modellen mit Hilfe von XAI adressieren lässt, sind dem Einsatz entsprechender Methoden auch klare Grenzen gesetzt. Selbst die am weitesten verbreiteten XAI-Ansätze können nicht für sämtliche KI-Modelle herangezogen werden. Denn diese beziehen sich in erster Linie auf klassische Klassifikationsmodelle.

Ein Beispiel für nicht betrachtete Modelle sind sogenannte Autoencoder. Deren Aufgabe ist es, verschiedene Eingabedaten zunächst in eine modellinterne, verdichtete Repräsentation zu überführen und diese im Anschluss wieder so nah wie möglich an die Eingabedaten zu transformieren. Solche Modelle können bspw. zur Anomalieerkennung (vgl. Kapitel 2.6) herangezogen werden. Sollen diese Modelle mithilfe von XAI erklärt werden, muss dies indirekt und über Umwege geschehen. Die Autoren Gnos, Schultz, und Tropmann-Frick zeigen dazu eine mögliche Vorgehensweise, indem die Vorhersagen des Autoencoders genutzt werden, um darauf basierend beschriftete Daten abzuleiten [24]. Diese beschriftete Datenbasis wird genutzt, um ein „traditionelles“ Klassifikationsmodell zu trainieren, welches dann

wiederum von den genannten XAI-Methoden erklärbar gemacht werden kann.
Das Resultat: ein erheblicher Mehraufwand und keine unmittelbare Erklärung des Ausgangsmodells.

Ebenfalls erwähnt werden muss, dass die Aussagekraft der von XAI-Ansätzen generierten Erklärungen nicht in dem Maße ausfällt, wie es für eine nachvollziehbare Darstellung des Modellverhaltens notwendig wäre. Dadurch kann die für Endanwender erforderliche, vollständige Transparenz und Nachvollziehbarkeit nicht immer sichergestellt werden.

2.6 Anomalieerkennung

Kurzvorstellung

Die Erkennung von Anomalien bezieht sich auf den Prozess der Identifizierung ungewöhnlicher Muster, Diskrepanzen oder verdächtiger Aktivitäten, die erheblich vom erwarteten bzw. normalen Verhalten innerhalb eines Datensatzes abweichen. Anomalieerkennung ein wichtiges Instrument in den verschiedensten Anwendungsbereichen, wie z. B. Wirtschaftsprüfung, Interne Revision, Cybersicherheit, Betrugserkennung und cyber-physische Systeme. Die Algorithmen zur Erkennung von Anomalien lassen sich in drei Kategorien unterteilen:

- (1) Statistische Methoden:** Die allgemeine Funktionsweise dieser Methoden basiert auf der Annahme, dass die Daten einer bekannten statistischen Verteilung folgen. So nutzt man mathematische Modelle wie die Gaußsche Verteilung oder die Zeitreihenanalyse, um Datenpunkte zu identifizieren, die erheblich von den erwarteten Mustern abweichen. Um einen Schwellenwert für die Identifizierung von Anomalien festzulegen, werden bei diesen Methoden häufig statistische Metriken wie Mittelwert, Standardabweichung oder Quantile berechnet.
- (2) ML-basierte Methoden:** Auf maschinellem Lernen basierende Methoden werden häufig zur Erkennung von Anomalien eingesetzt. Dabei kommen überwachte Lernalgorithmen zum Einsatz, die auf markierten Daten trainiert werden, um Anomalien explizit identifizieren zu können. Zugleich werden auf der Grundlage der erlernten Muster neue Datenpunkte als normal oder anomal definiert. Demgegenüber lernen unüberwachte Lernalgorithmen, wie z. B. Clustering oder dichte-basierte Methoden, anhand der vorhandenen Daten normale Muster zu erkennen und kennzeichnen Datenpunkte als Anomalien, die diesen Mustern nicht entsprechen.
- (3) Regel-basierte Methoden:** Bei regelbasierten Ansätzen wird eine Reihe von vordefinierten Regeln oder Heuristiken definiert, die das erwartete Verhalten des Systems oder des Datensatzes erfassen. Die eingesetzten Regeln beruhen in der Regel auf Expertenwissen oder domänenspezifischen Richtlinien, die für ein spezifisches Problemfeld entworfen und implementiert wurden. Bei der Erkennung von Anomalien mit Hilfe regelbasierter Methoden werden Datenpunkte anhand dieser vordefinierten Regeln überprüft, um Verstöße oder Abweichungen festzustellen. Verstößt ein Datenpunkt gegen eine oder mehrere Regeln, wird dieser als Anomalie gekennzeichnet. Regelbasierte Ansätze sind oft transparent und interpretierbar, da die Bedingungen für die Erkennung von Anomalien seitens der Regeln explizit beschrieben sind.

Für optimale Ergebnisse ist es ratsam, je nach Anwendungsfall ggf. verschiedene Verfahren zur Anomalieerkennung zu kombinieren. Auf diese Weise lassen sich die von jeder einzelnen Methode mitgebrachten Vorteile bestmöglich ausnutzen.

Bedeutung und Nutzen für den Steuerbereich

Aufgrund der komplexen und dynamischen Natur steuerbezogener Daten, bieten Techniken zur Erkennung von Anomalien ein immenses Potenzial für den Steuerbereich. Allein die Analyse verschiedener, steuerbezogener Datensätze (z. B. Finanztransaktionen, Gewinn- und Verlustrechnungen oder Steuererklärungen) kann dazu beitragen, potenziellen Steuerbetrug, Fehler oder Unregelmäßigkeiten frühzeitig zu erkennen. Sie identifiziert Vergehen oder Fehler, die sonst möglicherweise unbemerkt geblieben und zu Geldstrafen und Reputationsverlusten geführt hätten.

- (1) Effiziente Steuerprüfungen:** Während die manuelle Überprüfung von Finanzdaten recht zeitaufwendig ist, können automatisierte Verfahren zur Erkennung von Anomalien den Prozess beschleunigen. Indem automatisch ungewöhnliche Muster identifiziert und frühzeitig auf potenzielle Risiken hinweisen, kann insgesamt die Wirtschaftlichkeit des Prüfprozesses verbessert werden. Außerdem können sich dadurch die Steuerfachleute auf Fälle mit hoher Priorität (z. B. hohes Steuerrisiko) oder auf wissensintensive Aufgaben (z. B. strategische Steuerplanung) fokussieren, wodurch die Qualität der Prüfung insgesamt gesteigert werden kann.
- (2) Steuerliche Compliance:** Steuervorschriften sind nicht selten sehr komplex und können sich im Laufe der Zeit ändern. Bei Bedarf lassen sich die im System hinterlegten Steuerregeln mit Hilfe der Anomalieerkennung aber relativ schnell anpassen. Dadurch wird sichergestellt, dass Unternehmen mit den neuesten steuerlichen Vorschriften konform sind.

Grenzen für die Nutzung im Steuerbereich

Eine der größten Herausforderungen bei der Erkennung von Anomalien ist das Risiko von Fehlalarmen. Ein System könnte z. B. eine Transaktion als Anomalie erkennen, obwohl diese eigentlich normal ist (false positive) oder eine anomale Transaktion als normal kennzeichnen (false negative). Aufgrund falsch positiver bzw. falsch negativer Ergebnisse kann es in der Folge zu Ineffizienzen bei der Prüfung kommen. Im schlimmsten Fall werden tatsächliche Anomalien komplett übersehen.

Bei ML-basierten sowie statischen Methoden hängt die Effizienz und Effektivität des Modells zudem stark von der Qualität der Daten ab. So können bspw. ungenaue oder unvollständige Datensätze zu irreführenden Ergebnissen führen. Ferner ist die Vorhersage stark abhängig von der Qualität der verwendeten Merkmale (Features) und der Dimensionalität der Daten, also wie viele Referenzinformationen zu einem messbaren Ereignis vorliegen und im Modelltraining mitberücksichtigt werden. Eine weitere Herausforderung stellt die Veränderung der Daten dar. Algorithmen zur Anomalieerkennung können sich nur schwer an neue Arten von Anomalien oder sich ändernde Daten anpassen. So kann es für das Modell schwierig sein, zwischen echten Anomalien und legitimen Verschiebungen in der zugrundeliegenden Datenverteilung zu unterscheiden.

2.7 Klassifikationsmethoden

Kurzvorstellung

Bei der Klassifizierung handelt es sich um eine Art des Überwachten Lernens. Ziel dieser Methode ist es, die kategorialen Klassenbezeichnungen neuer Datenpunkte vorherzusagen – und zwar auf der Grundlage früherer Beobachtungen. Oder anders formuliert: Ein Klassifizierungsalgorithmus lernt aus einem Datensatz mit bekannten Klassenbezeichnungen die Zuordnung neuer Datenpunkte in eine oder mehreren Klassen.

Bei vielen realen und betriebswirtschaftlichen Problemstellungen handelt es sich im Wesentlichen um Klassifizierungsprobleme. Dadurch lässt sich die Klassifikation auch als grundlegendes Konzept interpretieren, das nicht nur für eine Vielzahl praktischer Anwendungen genutzt wird. Es dient auch als Basis für die Entwicklung und Evaluation komplexer maschineller Lernsysteme

Bedeutung und Nutzen für den Steuerbereich

Gerade im Steuerbereich können Algorithmen zur Klassifikation die Steuerfunktion maßgeblich unterstützen. Denn jede Transaktion, jeder Vermögenswert und jeder Abzug wirkt sich unmittelbar auf die steuerliche Last eines Unternehmens aus. Je nach Aufgabe kann dabei zwischen drei Arten der Klassifikation unterschieden werden: Während sich die binäre Klassifikation durch genau zwei Klassenlabels auszeichnet, können bei der Mehrklassen-Klassifikation auch mehr als zwei Klassenlabels vorhanden sein. Demgegenüber bezeichnet die Multi-Label-Klassifikation ein Klassifikationsproblem, bei dem neue Daten einem oder gleich mehreren Klassen zugeordnet werden können.

- (1) Binäre Klassifikation:** Innerhalb eines Geschäftsjahrs entstehen eine Vielzahl unterschiedlicher Betriebsausgaben, die allerdings nicht alle in gleicher Höhe bzw. überhaupt steuerlich absetzbar sind. Umso herausfordernder ist es für Unternehmen, genau jene Ausgaben zu identifizieren, die tatsächlich absetzbar sind. Durch entsprechende Vorhersagen kann ein binäres Klassifizierungsmodell den Steuerexperten dabei unterstützen, herauszufinden, inwiefern bestimmte Geschäftsausgaben steuerlich absetzbar sind oder nicht.
- (2) Mehrklassen-Klassifikation:** Eine Vielzahl gesetzlicher und organisatorischer Gegebenheiten beeinflussen die Bestimmung der Umsatzsteuer. Für Unternehmen und insbesondere für solche, die unterschiedlichen Produktportfolios oder Dienstleistungen tätig sind, kann die Zuordnung des richtigen Steuersatzes eine echte Herausforderung werden. Unterstützung erhält der Steuerexperte seitens des Mehrklassen-Klassifizierungsmodells, das lernt jeder Art von Transaktion den richtigen Steuersatz zuzuordnen.
- (3) Multi-Label-Klassifikation:** Werden Güter im Ausland eingekauft, kann eine Transaktion mehrere steuerliche Auswirkungen gleichzeitig haben. So kann ein einziger Vorgang bei der Bearbeitung beispielsweise als „einfuhrzollpflichtig“ und „mehrwertsteuerpflichtig“ gekennzeichnet sein. Im Gegensatz zur Mehrklassen-Klassifikation, die aus mehreren Optionen eine Kategorie zuordnen würde, unterbreitet die Multi-Label-Klassifikation dem Steuerexperten neben einem Vorschlag auch die Zuordnung gleich mehrerer Kategorien.

Grenzen für die Nutzung im Steuerbereich

Die spezifische Auswahl der verwendeten Algorithmen und Strategien hängen immer auch von den Grenzen eines KI-Systems ab. So sind Klassifikationsalgorithmen stark von der Qualität und Quantität der Daten abhängig, auf denen sie trainiert wurden. Sind diese verzerrt, unvollständig, unausgewogen oder nicht repräsentativ, können die Vorhersagen des Modells ungenau oder schlimmstenfalls fehlerhaft sein.

Beim Einsatz von Algorithmen kann es ebenso zu Schwierigkeiten kommen, wenn die Anzahl der Datenattribute (Merkmale) im Vergleich zur Anzahl der Trainingsinstanzen zu hoch ist. Ein weiterer wichtiger Aspekt ist die Interpretierbarkeit des Modells. Einige Klassifizierungsmodelle wie z. B. tiefe

neuronale Netze, werden oft als Blackbox-Modelle betrachtet, denen es an Interpretierbarkeit fehlt. Entsprechend schwierig ist es auch zu verstehen, wie das Modell zu bestimmten Entscheidungen kommt. Ein Umstand, der besonders in jenen Bereichen problematisch ist, wo maximale Transparenz erforderlich ist.

Wird der Klassifikationsalgorithmus zu eng auf die Trainingsdaten zugeschnitten, lässt sich das Modell unter Umständen nicht auf neue Daten verallgemeinern. Ähnliches gilt, wenn das Modell zu kurz oder auf zu wenig Daten trainiert wurde. Zudem können Klassifizierungsmodelle, die auf historischen Daten trainiert wurden, im Laufe der Zeit an Genauigkeit verlieren. Ein Problem, das immer genau dann auftritt, wenn sich die zugrundeliegenden Muster und Beziehungen in den Daten ändern sollten. Das kann sich insbesondere im Steuerbereich negativ auf die Leistung des verwendeten Klassifikationsmodells auswirken, da die Werte in den Datensätzen aufgrund neuer Vorschriften, Steuergesetze und wirtschaftlicher Veränderungen mitunter stark schwanken können.

2.8 Synthetische Daten

Kurzvorstellung

Im Gegensatz zu einem modellzentrierten Ansatz wird mit einem datenzentrierten Ansatz (engl. data-centric AI) das Ziel verfolgt, KI-Systeme auf Grundlage qualitativ hochwertiger und repräsentativer Daten zu entwickeln. Eine Möglichkeit dieses Ziel zu erreichen, beschreibt die Generierung synthetischer Daten [25], [26]. Dabei handelt es sich um Daten, die mit Hilfe eines mathematischen Modells oder Algorithmus erzeugt werden und deren statistischen Merkmale erhalten bleiben. Synthetische Daten sind für Unternehmen aus zweierlei Gründen interessant: Training und Verbesserung von KI-Modellen sowie Datenschutz [27].

Mit entsprechende Ansätzen verbessern bspw. bereits Waymo [28] und Tesla [29] ihre KI-basierten Fahrerassistenzsysteme. Da synthetische Daten im Allgemeinen keine sensiblen oder persönlichen Informationen enthalten, kann auch ein gewisser Schutz der Privatsphäre sichergestellt werden. Dadurch wird nicht nur die Verarbeitung von Daten z. B. in Kooperation mit externen Dienstleistern ermöglicht. Auch die versehentliche Weitergabe von sensiblen Informationen (z. B. durch *Model Inversion* oder *Membership Inference Attacks*) lässt sich so verhindern.

Bedeutung und Nutzen für den Steuerbereich

Für den Steuerbereich versprechen Ansätze und Techniken zur Generierung synthetischer Daten neue Möglichkeiten. So kann nicht nur die Datenvielfalt gezielt erweitert werden, indem Szenarien erzeugt werden, die in den ursprünglichen Daten nicht enthalten waren. Aber auch seltene Ereignisse lassen sich simulieren, sodass das KI-Modell in der Lage ist, diese zu erkennen und darauf zu reagieren. Außerdem lässt sich vor allem die Wahrscheinlichkeit einer Überanpassung (engl. Overfitting) und einer möglichen Preisgabe sensibler Datenpunkte aus dem KI-Modell auf ein Minimum reduzieren. Ein Effekt, den nicht zuletzt eine Studie vom MIT-IBM Watson AI Lab und der Boston University offenbarte. Demnach konnten anhand des Trainings von Deep-Learning-basierten Modellen die Überanpassung verringert und die Genauigkeit insgesamt verbessert werden [30].

Aber auch das Problem unausgeglichener Daten, die zu einer großen Anzahl von Fehlalarmen führen und somit für Unternehmen kostspielig sein können, kann durch synthetische Daten verringert werden. Hierbei können Datenpunkte der Minderheitsklasse vermehrt werden. In diesem Zusammenhang können synthetische Daten auch zur Lösung von Problemen bezüglich des Model- und Concept-Drifts beitragen, insbesondere dann, wenn die Erhebung neuer realer Daten schwierig oder teuer ist.

Hierunter versteht man die Veränderung von Annahmen oder Variablenabhängigkeiten bzw. der statistischen Eigenschaften der Ziele, wodurch die Vorhersagegenauigkeit von KI-Modellen mit der Zeit nachlassen kann.

Zudem unterstützen synthetische Daten die Entwicklung genauerer ML-Modelle, ohne dabei gegen datenschutzrechtliche Bestimmungen zu verstoßen. Fehlen etwa Personal- oder Hardwareressourcen, lassen sich unkompliziert externe Dienstleister bzw. Cloud-Plattformen in den Entwicklungsprozess einbinden bzw. nutzen. Ferner können durch Simulationen eine Vielzahl hypothetischer Steuerszenarien generiert werden, die bei Prognosen und steuerpolitischen Entscheidungen hilfreich sein können. Auswirkungen auf neue Steuerreformen oder Änderungen der Steuergesetzgebung lassen sich dadurch relativ schnell bewerten. Ebenso können aufwendige und kostenintensive Verfahren zur Erfassung „echter“ Daten auf ein Minimum reduziert werden.

Einige KI-Forscher warnen vor „irreversiblen Defekten“ durch KI-generierte Datensätze [31]. Dennoch ist zunehmend ein Trend hin zu datenzentrierten Ansätzen und hin zur Generierung durch große Sprachmodelle wie ChatGPT oder Deep-Learning-Modelle wie CTGAN bzw. VAE erkennbar. Im Fokus steht dabei nicht die Verbesserung der KI-Modelle und deren Hyperparameter, sondern die Bereitstellung qualitativ hochwertiger Trainingsdaten. Auch Sam Altman, CEO von OpenAI, zeigt sich zuversichtlich, dass bald alle Daten für das Training großer KI-Modelle synthetisch sein werden. Das könnte insbesondere OpenAI dabei helfen, die Datenschutzbedenken in der EU zu umgehen. Aber auch Aiden Gomez, CEO des LLM-Startups Cohere, ist davon überzeugt, dass die aktuellen Herausforderungen im Umgang mit Webdaten durch synthetische Daten gelöst werden können [32]. Denn dadurch lässt sich das Problem von Webdaten beseitigen, die oft verrauscht oder verzerrt und somit nicht wirklich repräsentativ für das Training von ML-Modellen sind.

Je nach Anwendungskontext und Art der Daten (z. B. in Form von Text, tabellarischen Daten oder Bildern) lassen sich die Ansätze grundsätzlich in zwei Kategorien unterteilen:

- (1) Synthese aus realen Daten:** Hierbei werden aus realen Datensätzen neue Datenpunkte synthetisiert. Der Datenanalytiker erstellt anhand eines realen Datensatzes ein Modell, das die Verteilungen und die Struktur der realen Daten erfasst. Je nach Anwendungsfall können dabei unterschiedliche Techniken bzw. Modell-Architekturen zum Einsatz kommen.
- (2) Synthese ohne reale Daten:** In diesem Fall werden die synthetischen Daten nicht aus realen Datensätzen generiert, sondern unter Verwendung vorhandener Modelle oder des Fachwissens bzw. den Erfahrungen des Fach- und Datenexperten. Dabei kann es sich um regelbasierte Verfahren, statistische Modelle (basierend auf Umfragen oder andere Mechanismen zur Datenerfassung) oder Simulationen handeln.

Grenzen für die Nutzung im Steuerbereich

Auch wenn die Verwendung synthetischer Daten eine Vielzahl von Vorteilen mit sich bringt, ist sie mit einigen Herausforderungen verbunden. Beispielsweise kann es aufgrund der inhärenten Strukturen in den realen Daten zu Ungenauigkeiten oder Verzerrungen in den synthetischen Daten kommen. Anders formuliert: Die synthetischen Daten sind immer nur so gut wie das Modell selbst, mit dem diese erstellt wurden. Genau deshalb braucht es auch entsprechende Qualitäts- und Validierungsmaßnahmen, um die erforderliche Genauigkeit, Zuverlässigkeit und Privatsphäre der Daten sicherstellen zu können.

Ein häufig zu beobachtendes Missverständnis bei synthetischen Daten ist, dass diese automatisch als „privat“ angesehen werden. Dies ist jedoch nicht immer der Fall, insbesondere dann, wenn diese aus anderen realen Daten abgeleitet werden und dadurch Rückschlüsse über diese ermöglichen [33]. Zudem kann es aufgrund von Ressourcenmangel zu Einschränkungen kommen, da es für die Generierung synthetischer Daten einen interdisziplinären Ansatz braucht. Damit gemeint sind verschiedene Methoden und Modelle, die fach- und bereichsübergreifend zum Einsatz kommen. Schließlich kann die Verwendung synthetischer Daten mit rechtlichen und ethischen Überlegungen verbunden sein. Das gilt insbesondere dann, wenn die Daten nicht vollständig verstanden werden oder es aufgrund fehlerhafter Sachverhalte zu steuerrechtlichen Verstößen durch das mit den synthetischen Daten erweiterte KI-Modell kommt.

3 KI-Anwendungen im Steuerbereich – Ergebnisse der Experteninterviews

3.1 Einführung

Die Entwicklung der KI schreitet mit beispielloser Geschwindigkeit voran und revolutioniert altbewährte Arbeitsprozesse in den unterschiedlichsten Branchen. Auch Steuerabteilungen sind davon nicht ausgenommen und sind zunehmend daran interessiert, neue KI-Technologien einzusetzen. Das zeigen auch verschiedene Studien der OECD [34] und von McKinsey [35] über den Einfluss von KI auf die Steuer- und Arbeitswelt, wonach die Integration von KI die Arbeitsweise von Fachleuten in diesen Bereichen nachhaltig verändern kann. Heißt konkret: Auf kurz oder lang werden KI-Technologien und Modelle einen entscheidenden Beitrag dazu leisten, Arbeitsprozesse effektiver und effizienter zu gestalten und sowohl Entscheidungsfindungen als auch die Steuerkonformität zu verbessern.

Technologien wie Natural Language Processing und Maschinelles Lernen sind vor allem dann geeignet, wenn sich wiederholende und alltägliche Aufgaben bzw. Aktivitäten in Steuerprozessen automatisiert werden sollen. Allein die Kombination von OCR und KI kann bspw. dabei helfen, die Erstellung von Steuererklärungen, Vertragsanalysen und Rechtsrecherchen zu „computerisieren“. Dadurch können sich Steuerexperten auf weitaus komplexere und strategisch wichtigere Aufgaben konzentrieren. In diesem Zusammenhang beschreibt Zergibel zum Beispiel, wie sich mit Hilfe von KI-Technologien sowohl Eingangs- als auch Ausgangsrechnungen automatisiert erfassen und digital aufbereiten lassen [36]. Dabei lernt ein KI-System, mandantenspezifisch die Struktur der Rechnungen zu erfassen, um entsprechende Belege teil- oder vollautomatisch zu generieren bzw. zu verbuchen.

Darüber hinaus kommen ML-Algorithmen immer dann zum Einsatz, wenn große Datenmengen wie z.B. steuerrelevante Transaktionsdaten nach bestimmten Mustern und Trends durchsucht werden sollen. Dadurch werden Steuerexperten darin befähigt, innerhalb kürzerer Zeit fundiertere Entscheidungen zu treffen. Aber nicht nur Auffälligkeiten können effizient identifiziert werden. Durch die zusätzliche Berücksichtigung weiterer Faktoren lassen sich auch Empfehlungen für Steuerplanungs- und Optimierungsstrategien generieren. Selbst bei der Recherche in Rechtsdatenbanken oder Gesetzestexten können KI-Systeme wertvolle Hilfe leisten und den Steuerexperten maßgeblich bei seiner Arbeit unterstützen – beispielsweise durch das Extrahieren und Verdichten relevanter Informationen.

Durch den Einsatz von KI-Anwendungen im Steuerbereich lassen sich bspw. neuralgische Fallkonstellationen schneller identifizieren und vergleichbare Sachverhalte leichter finden, um dem Steuerexperten entsprechende Prüfungs- und Risikohinweise zur Verfügung zu stellen [37]. Auch Fragen zu steuerlichen Themen wie der Zusammenfassenden Meldung, der Steuernummer oder zollrelevante Informationen zum grenzüberschreitenden Waren-, Personen- und Dienstleistungsverkehr können bereitgestellt werden. Dies sind wichtige Aufgaben, die mittlerweile digitale Assistenten wie zum Beispiel *ViOIA* [38] oder *TinA* [39] übernehmen.

Im Folgenden werden zentrale KI-Anwendungen und ihre Potenziale für den Steuerbereich vorgestellt, die entweder bereits zum Einsatz kommen oder zumindest von hoher Relevanz sind. Ziel ist es, nicht nur die damit verbundenen Chancen und Herausforderungen darzustellen. Auch der Einfluss von KI-Anwendungen auf die verschiedensten Arbeitsprozesse soll anhand von Experteninterviews näher beleuchtet werden, die innerhalb der WTS durchgeführt wurden.

3.2 Anwendungen im Bereich HR Taxes

Allgemeine Informationen

Löhne, Zuschläge, Abfindungen, Geschenke, Mahlzeiten, Dienstreisen, (Reisekosten-) Erstattungen, sonstige Sachzuwendungen, geldwerte Vorteile, Veranstaltungen, Dienstfahräder, Dienstwagen, Entsendungen, Work from anywhere, Home-Office (im Ausland) – all das sind Themen, die beim Arbeitgeber zu Prozess- und Steuerfragen der unterschiedlichsten Art und Komplexität führen.

Die damit verbundenen Prozesse liegen in der Regel über das gesamte Unternehmen verteilt, weshalb sich sehr viele verschiedene Funktionen mit deren Bearbeitung befassen. Der Schwerpunkt liegt in diesem Zusammenhang aber ganz klar im Personal- und Rechnungswesen, wobei in der Regel die Personalfunktion (haupt-)verantwortlich ist. Die Steuerabteilung besitzt zumeist „nur“ eine unterstützende Funktion und übernimmt seit der starken Zunahme von Tax CMS auch eine kontrollierende Rolle.

Früher sprach man hier „einfach“ von „Lohnsteuer“. Heutzutage wird der Bereich zutreffender mit „HR-Taxes“ bezeichnet. Ein Begriff, der die gesamte Bandbreite der Steuerfragen im Bereich der personen- und arbeitnehmerbezogenen Steuern ausdrücken soll, denn „Lohnsteuer“ ist hiervon nur ein kleiner Teil.

HR Taxes umfassen auch vielfältige, zum Teil pauschale Steuererhebungsformen (in § 40 EStG, §§ 40a, 40b EStG, die pauschale Einkommensteuer auf Sachzuwendungen nach § 37b EStG). Weiter zwingend ist das Zusammenspiel mit den Sozialversicherungsabgaben, der Umsatzsteuer, den Verrechnungspreisen, Betriebstätten Themen sowie Themen der Einkommensteuer bei Steuererklärungen (bei nationalen & grenzüberschreitenden Sachverhalten). Große Bedeutung besitzt auch der Bereich der sogenannten Shadow Payroll. Der Schlüssel ist daher die ganzheitliche Betrachtung und Abstimmung aller steuerlichen Sachverhalte bei Befassung mit HR-Taxes.

Immer relevanter werden im Bereich der Sachzuwendungen auch die Zusammenhänge mit den immer strikter werdenden allgemeinen Compliance-Regeln der Unternehmen. Ebenfalls rücken mit Blick auf die hohen Kosten Aspekte wie die kosten-, steuer- und sozialversicherungsorientierte Planung von Sachzuwendungen und Veranstaltungen in den Fokus. In der Folge kommt es oft zu einer Einschränkung und Standardisierung von Zuwendungsarten.

In großen Unternehmen sind oft tausende von Mitarbeitern mit HR-Taxes befasst. Die allermeisten verfügen dabei allerdings über keinerlei steuerlichen Background. Aus diesem Grund geht heute die klare Tendenz bei der Ausgestaltung von HR-Tax Prozessen weg von Schulungen und (schriftlichen) Arbeitsanweisungen (die oft immer weniger gelesen und verstanden werden) hin zu geführten Prozessen mit Entscheidungsbäumen. Diese werden in bestehende Systeme eingebettet und sollen zukünftig zu einer möglichst vollständigen Digitalisierung der Steuer führen. Nur über ständig zu verbessernde Arbeitsprozesse kann eine hinreichende Kosten- und Steuereffizienz oder besser die Compliance in der Zukunft sichergestellt werden.

Wichtig ist hier, dass der geführte und später der digitale Prozess, soweit möglich, alle der folgenden, bislang manuell abgebildeten Teilschritte übernimmt:

- (1) Erkennen
- (2) Erfassen
- (3) Berechnen
- (4) Validieren
- (5) Export / Handling

Experteninterview

mit Christiane Noatsch und Philipp Besson

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

*Was konkret seht ihr, wenn ihr euch die Steuerfunktion der Zukunft im Unternehmen vorstellt?
Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?*

Gerade im Bereich Lohnsteuer arbeiten wir mit sehr großen Datenmengen. Durch die Kombination von steuerlichem Fachwissen und neuen Technologien lassen sich gerade hier viele Synergien nutzen. Um jedoch fachliche Inhalte zu validieren und passende Prompts zu erstellen, braucht es erfahrene Kollegen. Aber auch neue Kollegen und Bewerber sind gefragt, die unbelastete Fragen stellen und eine gewisse Affinität für Technologien und Prozesse mitbringen. Mitarbeiter, die sich nicht scheuen, im Morgen zu leben und zu denken.

Speziell im Bereich der Prozesssteuern (Umsatzsteuer, Lohnsteuer, etc.) ist das Thema KI „DIE“ große Chance. Denn gerade hier war man in der Vergangenheit technologieeitig oft von der Steuerabteilung abgeschnitten. Obwohl Prozesssteuern als dominierende Risikosteuern eines Unternehmens betrachtet werden müssen, wurde die Umsatzsteuer vom Accounting als „Abfallprodukt“ bewertet und die HR-Taxes über das Personalwesen (HR) bearbeitet. Hier kann oder besser muss jetzt ein Wandel angestoßen werden.

Zudem ist das Thema Steuern durch Skandale und Steuervermeidungstaktiken oft negativ behaftet. Hier muss sich die Wahrnehmung ändern vom „Verhinderer“ zum „Mitgestalter“ – aufbauend auf den Unternehmensprozessen. Wir müssen hin zu „Business drives Tax“. Doch dafür braucht man Leute, die prozessaffin sind und über ein gewisses technisches Verständnis bzw. steuerliche Grundlagen verfügen. Der zukünftige Schwerpunkt wird klar auf den Prozessen liegen, sowie deren Kontrolle und der Fortentwicklung. Das Ziel: 80-90% der Sachverhalte werden über geführte Prozesse und KI abgebildet. Die restlichen 10% werden weiterhin manuell von qualitativ hochwertigen Fachkräften betreut.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

Im Bereich der HR-Taxes kam KI praktisch kaum zum Einsatz. Am ehesten wurde hier in den letzten Jahren auf die Nutzung von OCR-Lösungen zurückgegriffen, um PDF-Dokumente auszulesen. Der Bedarf für solche KI-Tools in unserem Bereich ist relativ hoch, konnten damit in den vergangenen Jahren allerdings noch keinen zählbaren Mehrwert generieren.

In diesem Zusammenhang kommt immer wieder auch die Frage auf: Was ist unter KI eigentlich zu verstehen? Was in der Vergangenheit unter KI verstanden wurde, konnten wir mit Hilfe von Entscheidungsbäumen und geführten Prozessen umsetzen. Diese Projekte waren erfolgreich, sofern die Komplexität geringgehalten werden konnte. Demgegenüber haben komplexe Lösungen oft nur in der Theorie und nicht in der Praxis funktioniert. Rückblickend lässt sich behaupten, dass unser Verständnis von KI im Jahr 2017 in Wahrheit nichts mit einer richtigen KI zu tun hatte.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Insbesondere im Bereich der HR-Taxes ist das Potenzial für KI enorm hoch. Das zeigt auch das Interesse unserer Mandanten an technologischen Lösungen für §37b EStG Sachverhalte, die es tatsächlich dringend braucht. Allerdings ist die Komplexität in diesem Bereich darin begründet, dass HR-Taxes sehr viele Schnittstellen zu anderen Rechtsgebieten hat. So müssen bestehende Prozesse vorab erst intensiv analysiert werden, um das eigentlich vorhandene Potenzial voll ausschöpfen zu können.

Eine zentrale Voraussetzung für die Nutzung von KI ist, dass Unternehmen in Zukunft umschwenken und umdenken – von einer reaktiven hin zu einer proaktiven Bearbeitung von Sachverhalten. Hier zeigt auch das TaxCMS, dass solche Prozesse benötigt werden. Denn sonst drohen bei intensiven Lohnsteuerprüfungen hohe Anpassungen.

Grundsätzlich ergibt sich bei HR-Taxes eine enorme Möglichkeit für den Einsatz von KI. Insbesondere was die Lohnsteuer und Sachzuwendungsprozesse als Massenprozess betrifft. Hier braucht es aber eine Art „geführte KI“, da das Thema nicht ausschließlich anhand vorliegender Rechtsquellen bearbeitet werden kann und auch den nötigen Spielraum für Ermessen lässt. Soll heißen: Wenn Unternehmen sich eigene Risikoprofile definieren, muss der KI auch eine gewisse Guidance an die Hand gegeben werden.

So viel steht jedenfalls fest: KI-Projekte scheitern in der Regel nicht am Budget. Oft liegt es eher an der mangelnden Bereitschaft, etwas zu verändern, oder am „Tone from the Top“. Auf der anderen Seite wird noch viel zu oft in Steuer-Silos gedacht, anstatt sich Steuerarten übergreifend auszutauschen. Hier kann eine KI ebenfalls helfen. Denn anstatt nur im Silo zu denken, zeigt sie auch Schnittstellen zu anderen Steuer- und Rechtsgebieten auf oder nutzt diese sogar ganz gezielt.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Gerade für die Nutzung von generativen Sprachmodellen wie z. B. ChatGPT gibt es im HR-Taxes Bereich eine Vielzahl von inhaltlichen Anwendungsfällen. Zum Beispiel das Thema Mahlzeitengestellung und die Beantwortung der Frage: Wann wird hier Lohnsteuer fällig und wann sind nicht abziehbare Betriebsausgaben zu buchen etc.? Immer in Abhängig davon, ob es sich um eine Dienstreise handelt, eine Incentivierung, ein Arbeitsessen oder eine Betriebsveranstaltung.

Aber auch das Thema Gutscheine als Sachzuwendung eignet sich gut für eine KI-Lösung. Im Falle einer Rechtsberatung kann die KI diesbezüglich konkrete Unterstützung bieten. Oder man trainiert das Modell gezielt mit Richtlinien, BMF-Schreiben, Gesetzen und Rechtsprechungen. Immer unter Berücksichtigung des Zeitbezugs und möglicher Änderungen. Auch bei der Reisekostenabrechnung als Massensachverhalt kann eine KI zur Prüfung eingesetzt werden. Gerade hier kann die KI einen wertvollen Mehrwert generieren, da dieses Thema oft nur stiefmütterlich in Unternehmen behandelt wird.

Ein anderes großes Thema im Bereich HR-Taxes ist die steuerliche Behandlung von Veranstaltungen. Hier sind oft unzählige Personen zuständig, die kaum oder gar nicht über den eigentlich notwendigen Sachverstand verfügen. Ursache hierfür ist eine Vielzahl von Informationen, die aus unterschiedlichen Systemen kombiniert werden müssen und für die es nach den gesetzlichen Vorgaben bzw. anhand der Unternehmensrichtlinien einen Beurteilungsentwurf zu erstellen gilt. Ziel ist es, dass am Ende nur noch ein Review durch einen Mitarbeiter notwendig ist. Doch bevor hier überhaupt eine KI zum Einsatz kommen kann, muss bereits zuvor ein einfacher und geführter Prozess klar abgebildet sein.

Für die genannten Anwendungsfälle, bei denen Antwortentwürfe auf Standardfragen vorbereitet werden, ist das Thema „gleich ist gleich“ sehr wichtig. Wenn eine KI von heute auf morgen mehr lernt und dadurch eine andere Antwort gibt als gestern, kann dies zu Problemen führen. Hier muss sichergestellt sein, dass Gleiches tatsächlich auch gleich beantwortet wird. Zwar müssen in Perioden Änderungen möglich sein, aber bei agilen Änderungen der Antwort sollte man entsprechend vorsichtig sein.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Was die KI betrifft, dürfen ethische, moralische und haftungsrechtliche Themen nicht außer Acht gelassen werden. Genau deshalb müssen wir klare Werte im Umgang damit definieren und umsetzen. Aber auch der Datenschutz ist ein großes Thema, weil wir in der Regel sehr viel mit personenbezogenen Daten zu tun haben. Um die Einhaltung der Datenschutzbestimmung sicherzustellen, braucht es eine Zertifizierung von KI-Lösungen, die eine Konformität mit der DSGVO garantieren.

Abgesehen von den datenschutzrechtlichen Hürden braucht es aber auch die nötigen Skills, um mit KI-Tools die täglichen Herausforderungen erfolgreich bewältigen zu können. In diesem Zusammenhang müssen wir möglicherweise auch unsere Beratungsleistungen umstellen und uns mehr auf Prozesse und weniger auf High-End-Fachberatung konzentrieren. Selbstverständlich werden auch die Kosten immer eine wichtige Rolle spielen, da für KI-Lösungen nur dann Budget freigegeben wird, wenn durch deren Einsatz nachweislich ein Mehrwert entsteht.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Hier muss klar zwischen kleinen, mittleren und großen Unternehmen unterschieden werden. Gerade bei kleinen Unternehmen ist die fachliche Beurteilung aller Steuerthemen durch eine Person sehr schwer. Soll heißen: die Rolle und Aufgaben eines Beraters werden hier in Zukunft stärker geprägt sein durch Fähigkeiten, Ergebnisse gegebenenfalls mit KI zu erzeugen und dafür auch eine Haftung zu übernehmen.

Mittelgroße Unternehmen beschäftigen sich oft mit Spezialthemen, bei denen anspruchsvolle Steuerplanung anfängt und gewisse Prozessrisiken entstehen. Auch hier können diese Themen nicht ohne Weiteres von zwei oder drei Leuten abgedeckt werden. Heißt konkret: Selbst wenn eine KI zum Einsatz kommt, sollte immer ein Berater involviert sein.

Große Unternehmen sollten unbedingt Spezialisten in den eigenen Reihen wissen, die den Umgang mit technologischen Entwicklungen problemlos beherrschen. Experten, die KI verstehen und mit ihr umgehen können. Aber auch hier braucht es zwingend die fachliche Expertise, um die Beurteilung eines Beraters bewerten und hinterfragen zu können.

Zusammenfassend lässt sich sagen: Bei den Beratern im Steuerbereich muss unbedingt ein Umdenken stattfinden - gerade was die „älteren Generationen“ der Kollegen- und Kolleginnen betrifft. Hier wird die Agilität zur Normalität werden. Dementsprechend braucht es neue Gebührenmodelle, die nicht mehr auf Zeit basieren. Zugleich werden weiterhin aber auch Teamplayer mit Erfahrung und Expertise benötigt, die in der Lage sind die KI zu steuern und gemeinsam deren Ergebnisse zu beurteilen.

3.3 Anwendungen im Bereich Umsatzsteuer

Allgemeine Informationen

Bei der Umsatzsteuer (USt., englisch Value-Added Tax bzw. VAT) handelt es sich um eine indirekte Steuer auf die Wertschöpfung eines Unternehmens. Sie wird als Teil des Entgelts für Lieferungen und sonstige Leistungen eines Unternehmens erhoben. Bis auf wenige Ausnahmen sind davon alle angebotenen Produkte und Dienstleistungen betroffen. Auf vom Unternehmen zu begleichenden Rechnungen wird die ausgewiesene Umsatzsteuer auch als Vorsteuer bezeichnet. Rechtliche Vorgaben und Regelungen zur Umsatzbesteuerung sind im Umsatzsteuergesetz (UStG) definiert.

Für Unternehmen zählt die Umsatzsteuer nicht zu den betrieblichen Kosten, sondern bildet regelmäßig einen durchlaufenden Posten. Unternehmen müssen die Umsatzsteuer auf ihre angebotenen Produkte und Dienstleistungen in Eigenverantwortung direkt vom Kunden erheben und ans Finanzamt abführen. Aktuell erfolgt die Abrechnung mit dem Finanzamt im Rahmen der Umsatzsteuer-Voranmeldung und Umsatzsteuer-Jahreserklärung noch in einem regelmäßigen Turnus. Zukünftig ist eine zeitnahe Meldung auf Einzeltransaktionsbasis geplant. Im Falle einer nicht korrekt durchgeführten Umsatzsteuer-Abrechnung besteht die Gefahr, sich der Steuerhinterziehung schuldig zu machen.

Es existieren grundsätzlich zwei mögliche Vergehen:

- (1) Angabe von zu geringen Umsätzen
- (2) Zu hoch aufgeführte Vorsteuerbeträge

Zur Vermeidung von Umsatzsteuerbetrug ist es ein wesentliches Ziel der Steuerabteilung, die umsatzsteuerliche Voranmeldung bzw. die Meldung von Einzeltransaktionen möglichst korrekt durchzuführen. Insbesondere bei regelmäßig durchgeführten, hohen Korrekturen kann das Finanzamt ansonsten auch einen gewissen Vorsatz unterstellen.

Experteninterview

mit Gabriele Heemann und Nils Bleckmann

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret seht ihr, wenn ihr euch die Steuerfunktion der Zukunft im Unternehmen vorstellt? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Ich glaube, dass die Steuerabteilung der Zukunft den gleichen Auftrag haben wird wie heute. Hier zählen in erster Linie die Minimierung von Steuerrisiken und die optimale Erfüllung von Compliance Aufgaben. Die Einführung von Künstlicher Intelligenz beeinflusst dabei weniger die Aufgabe an sich. Vielmehr verändert sie die Art und Weise, wie Aufgaben erfüllt werden und welche Tools bzw. Werkzeuge dabei zum Einsatz kommen.

In dem Zusammenhang wird sich auch die Personalstruktur verändern. Waren früher in einer Steuerabteilung mit 60 Mitarbeitern vielleicht zwei bis drei Personen mit einem digitalen Auftrag betraut, werden es in Zukunft wohl eher zehn bis 15 oder sogar mehr sein. Eine Entwicklung, die in Zukunft vor allem durch die fortschreitende Digitalisierung des Steuerreportings vorangetrieben wird. Soll heißen, dass sich die Steuerabteilung funktional erheblich umfangreicher hin zum Daten-Owner und Daten-Processor wandelt.

Auch das Thema „Real Time Reporting“ wird den Bereich der Indirect Tax künftig nachhaltig prägen. Hier muss die Finanzverwaltung zeigen, dass die geforderten und erhobenen Datenmengen am Ende auch

tatsächlich verarbeitet werden können. Aktuell scheint das zumindest in Deutschland nicht der Fall zu sein, während andere Länder wie Italien oder Spanien hier bereits einen Schritt weiter sind.

Zusammengefasst wird es bei der Umsatzsteuer in der Zukunft nur noch um das Thema Daten gehen. Wer hat die Daten? Wer nutzt die Daten? Und wer bekommt diesen Prozess gesteuert?

Auch der klassische steuerliche Beratungsauftrag einer Steuerabteilung wird sich in Zukunft mit Veränderungen konfrontiert sehen: Wenn die großen Verlage ihre Informationen in generative Sprachmodelle wie bspw. ChatGPT überführen, dürfte sich vor allem der Research-Prozess wandeln. Die Steuerabteilungen werden dann vermehrt selber nach einer Lösung für ein Problem suchen, bevor sie einen Berater einschalten. Dieser tritt am Ende „nur“ noch in einer überprüfenden Rolle auf. Das alles hängt aber immer auch von der Größe eines Unternehmens ab. Kleineren Steuerabteilungen fehlt es in der Regel an den nötigen Kapazitäten, weshalb eine fachliche Durchdringung des Outputs einer KI weiterhin notwendig sein wird.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

KI wurde in den vergangenen Jahren hauptsächlich als Analyseelement für große Datenmengen verwendet. Es wurden Prüfroutinen geschaffen, die in aller Regel vergangenheitsorientiert waren. Hier beobachten wir aktuell einen Wandel, da die KI kreativer und auch zukunftsgerichtet wird. Allerdings stellt sich nach wie vor die Frage: Wo fängt KI an? Viele Unternehmen nutzen noch immer klassische Prüfroutinen und keine echte KI. Was jetzt möglich ist, kommt noch nicht einmal ansatzweise zum Einsatz.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Mit Blick auf die verschiedenen Einsatzmöglichkeiten wird deutlich, wie hoch das tatsächliche Potenzial der KI im Umsatzsteuerbereich doch ist. Hier können Sprachmodelle zum Beispiel im Rahmen der Rechnungsprüfung eingesetzt werden und einen Abgleich mit gesetzlichen Regelungen durchführen. Die automatisierte Verarbeitung von Texten ist eine der großen Stärken von Sprachmodellen, was auch ChatGPT mit seiner Einführung gezeigt hat. Doch in der Umsatzsteuer existieren im Vergleich mit anderen Steuerarten wie den Verrechnungspreisen weniger Dokumente mit viel Text. Stärker im Fokus stehen hier das Aufsetzen von Prozessen und die Massendatenverarbeitung.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Ein Anwendungsfall in der Umsatzsteuer ist zum Beispiel die Rechnungsprüfung. Daneben finden im Indirect Tax Bereich auch immer wieder Würdigungen von Einzelsachverhalten statt, bei denen es wichtig ist, eine Konsistenz in der Bewertung der Sachverhalte sicherzustellen. In der Vergangenheit wurde dies oft über Fallsammlungen in Steuerhandbüchern und Entscheidungsbäumen gemacht. Wenn man dieses Wissen allerdings in einen KI-basierten Chatbot einbringt, könnten bis zu 80% der Anfragen automatisiert beantwortet bzw. vorbereitet werden. Infolgedessen ließe sich auch die Konsistenz der Beantwortung sicherstellen.

Die KI ist vor allem auch für den Wissenstransfer sehr gut geeignet und könnte bspw. bei internen Schulungen zum Einsatz kommen. Aufgrund von Gesetzesänderungen und der Hinzunahme neuer Geschäftsvorfälle müssen vorhandene Schulungsunterlagen immer wieder angepasst werden. Mit Hilfe der KI lassen sich die Schulungsunterlagen bequem an die aktuelle Unternehmenssituation anpassen, um neue Mitarbeiter richtig abzuholen.

Der vielleicht größte Use Case für den Einsatz von KI hängt mit der aktuellen Entwicklung in Europa zusammen. Denn gerade im europäischen Wirtschaftsraum hat sich der Umsatzsteuerbereich hin zu einem Massendatenphänomen entwickelt. Im Unterschied zu den Ertragsteuern muss hier grundsätzlich jede Einzeltransaktion bewertet werden. Bisher hat das Finanzamt über die monatlichen Meldungen allerdings immer nur konsolidierte Informationen bekommen. Der aktuelle Trend in den Ländern geht mittlerweile hin zur Prüfung der Einzeltransaktionen im Rahmen eines „Real-Time-Meldewesens“, bei dem die Daten ohne Eingriffsmöglichkeit fortlaufend an das Finanzamt übermittelt werden müssen. Hier muss es der Technologie gelingen, innerhalb der Steuerabteilung ein Verständnis dafür zu vermitteln, was konkret ans Finanzamt gemeldet wurde und wo man für Einzelfälle vergleichbare Transaktionen findet. Hier liegt das wahre Steuerrisiko, weshalb die KI hier wirklich wertschöpfend sein kann.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Ein großes Risiko im Zusammenhang mit der KI sehe ich darin, dass sich ein bedingungsloser Glaube an die Maschine entwickeln könnte. Denn nicht selten wird angenommen, dass die KI immer alles richtig macht. Ganz im Gegenteil: Auch die KI macht Fehler, weshalb der Mensch auch weiterhin gebraucht wird – allein für die Einhaltung des 4-Augen Prinzips. Erfahrene Steuerexperten erkennen schnell, wenn die KI eine falsche Aussage macht. Da stellt sich nur die Frage: Erkennen das auch die jungen Kollegen aus einer Steuerabteilung?

Aber auch der Zugang zu KI ist eine Herausforderung: Große Unternehmen haben das Geld und die Manpower, um die von der KI mitgebrachten Vorteile gezielt zu nutzen. Aber was machen der Mittelstand und die kleinen Unternehmen ohne entsprechende Manpower und das nötige Kapital? Hier könnte in Zukunft eine Zweiklassengesellschaft im Steuerwesen drohen.

Und zu guter Letzt bleibt auch der Datenschutz ein sehr wichtiges Thema: Aktuell befinden sich die großen Sprachmodelle und die damit verbundenen Daten häufig auf Servern in den USA. Hier müssen vor allem die Mitarbeiter geschult werden, damit keine Firmengeheimnisse verraten werden.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Die Steuerabteilung alleine ist hier verloren, da sie noch nie eine IT-Abteilung war und bei den Mitarbeitern oft nur rudimentäre SAP-Kenntnisse vorhanden sind. Aus diesem Grund ist es noch wichtiger, dass sich die Steuerabteilung mit der IT-Abteilung und dem Accounting vernetzt. Der eigentliche Zweck von KI ist es ja, Ressourcen intelligenter einzusetzen und ggf. Personal einzusparen und nicht zehn neue Mitarbeiter aufzubauen, die KI nutzen können.

Die Berater hingegen müssen verstehen, dass die klassische Compliance allein nicht mehr den ausschlaggebenden Mehrwert für den Mandanten bietet. Vielmehr müssen sie den Nutzen der Automatisierung im Steuerbereich erklären können, dürfen dabei die Mitarbeiter im Unternehmen aber nicht verschrecken. Als Berater bleibt man nur dann im Rennen, wenn man zukunftsorientiert ist und



Lösungen als Intermediär vermittelt. Dazu kommen neue Anforderungen wie „VAT in the digital age (ViDA)“, die für Unternehmen ohne die Unterstützung von Beratern kaum zu bewältigen sind.

3.4 Anwendungen im Bereich Ertragsteuer

Allgemeine Informationen

Die Besteuerungsgrundlage für die Körperschaft- und Gewerbesteuer ist das Einkommen, das die Körperschaft innerhalb ihres Wirtschaftsjahres erzielt hat. Die Vorauszahlungen zur Körperschaft- und Gewerbesteuer werden im Rahmen der Veranlagung mit der tatsächlichen Steuerschuld verrechnet. Auf Unternehmensseite muss sichergestellt werden, dass stets genügend Liquidität zur Zahlung der Ertragsteuern verfügbar ist. Die korrekte Berechnung des zu versteuernden Gewinns und die sich daraus ableitende Belastung mit Körperschaft- und Gewerbesteuer gründet auf vielen gesetzlichen Vorschriften. Eine fehlerhafte Bestimmung der steuerlichen Belastung kann zu hohen Folgekosten und Liquiditätsengpässen führen.

Experteninterview

mit Dirk Beduhn, Dominik Lipp, Jochen Würges und Matthias Wulf

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret seht ihr, wenn ihr euch die Steuerfunktion der Zukunft im Unternehmen vorstellt? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Das Berufsbild in der Steuerfunktion wird sich in Zukunft verändern und in zwei Gruppen aufteilen: Zum einen in Fachspezialisten mit Berufserfahrung, die mehr Zeit mit steuerrechtlichen Themen verbringen und weniger manuellen Tätigkeiten nachgehen. Zum anderen in Mitarbeiter mit Technologie- und KI-Kompetenz, die sich hauptsächlich mit dem Management von Daten und der Formulierung von Prompts beschäftigen.

Auch heute schon fragen New Hires in Bewerbungsgesprächen, wie weit man im Digitalbereich ist. Denn gerade junge Leute denken weitaus digitaler als angenommen. Diese Kollegen müssen mit einbezogen werden, um Digitalisierungspotentiale zu erkennen und zu nutzen. Doch das gelingt nur, wenn ihnen auch die Freiheit gegeben wird, neue Technologien wie ChatGPT auszuprobieren und die eigene Kreativität zu fördern.

KI-Tools können vor allem einen großen Mehrwert und Zeitgewinn in der steuerlichen Basisarbeit liefern, zum Beispiel bei Literaturrecherchen oder der Beantwortung von steuerlichen Basisfragen. Dadurch bleibt

mehr Zeit für Beurteilungen auf Basis von materiellem Recht und komplexen Themen wie Umstrukturierungen. Gerade in diesem Bereich braucht es weiterhin Leute mit Erfahrung, die genau wissen wie man Sachverhalte aufarbeitet.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

Im Ertragsteuerbereich konnte man den Mehrwert von KI in den vergangenen Jahren teilweise im Einsatz von von Machine Learning und Chatbots mit generativen KI-Systemen beobachten. Allerdings war es hier eher schwierig, passende Anwendungsfälle zu finden, die skalier- und finanzierbar waren. Das mag auch daran liegen, dass der Steuerbereich im Unternehmen teilweise etwas „old-fashioned“ daherkommt und als Randbereich kein großes Investitionsvolumen besitzt. Dennoch hat man gesehen, dass sich die Prozesse im Unternehmen ändern und digitaler werden. Und obendrein auch der Prozessfokus zunimmt.

FRAGE 3: BLICK AUF DIE NAHE ZUKUNFT

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Die KI wird in naher Zukunft einen großen Vorteil gerade bei Recherchetätigkeiten bringen. Für die Beantwortung von Rechtsfragen geht es im ersten Schritt um die Beschaffung der „Rohstoffe“, also welche Rechtsprechung zur jeweiligen Frage existiert, welche Verwaltungsauffassung, welche Literatur und so weiter. Bei der manuellen Recherche besteht immer Ungewissheit, ob man alle relevanten Quellen gefunden hat. Demgegenüber lassen sich mit Hilfe der KI sowohl eine Qualitätssteigerung als auch eine Zeitersparnis erzielen.

Aber auch bei der Beantwortung klassischer Compliance-Fragen wird es durch den Einsatz von KI bequemer und schneller. So können Fragen zur Abzugsfähigkeit von Betriebsausgaben oder zu Abweichungen zwischen Handels- und Steuerbilanz beantwortet werden, ohne dass man dafür selbst entsprechende Richtlinien und Gesetze heraussuchen oder erst einen Berater für eine Auskunft kontaktieren muss.

Und auch beim Brainstorming liefern gerade die generativen Sprachmodelle wie ChatGPT große Vorteile. So hat unsere Studie gezeigt, dass diese Sprachmodelle häufig bei der Erstellung von Emails oder Textformulierungen verwendet werden. In erster Linie um Zeit zu sparen. Hat man Makros in Excel früher manuell und zeitaufwendig programmiert, kann ein generatives Sprachmodell diese Aufgabe heutzutage innerhalb weniger Sekunden erledigen.

Doch bei all den Vorteilen darf man auch die Nachteile nicht vergessen: Zum Beispiel wird das Lernen der Mitarbeiter erschwert, wenn kleine und einfache Aufgaben nur noch von der KI gelöst werden. Denn dadurch fehlt den Mitarbeitern am Ende die praktische Erfahrung bei solchen Aufgaben.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Im Accounting-Bereich geht es oft um die finale Berechnung der Steuern und die anschließende Erstellung des Anhangs. Hier braucht es Automatisierung und Standardisierung, um den entsprechenden fachlichen Input liefern zu können. Bei den Millionen von Einzelbelegen kann mit Hilfe von Künstlicher Intelligenz und Machine Learning eine Qualitätsverbesserung durch die Bereinigung der Datenquellen erzielt werden.

Auch bei der Erstellung des Endergebnisses kann die KI unterstützen. So könnte sie bspw. aus verschiedenen Datenquellen und unter Berücksichtigung der Regelungen des IAS12 einen konformen Anhang erstellen, der anschließend nur noch durch einen Mitarbeiter geprüft werden muss.

Gerade im Rahmen des Business Partnering ist das Thema Quellensteuer ein sehr wichtiges. Hier steht uns ein Fundus aus unzähligen Verträgen zur Verfügung – von Standardverträgen bis hin zu sehr komplexen Fällen. Wenn man dieses Wissen mit Hilfe der KI nutzbar macht, kann es sowohl bei der Beurteilung von Quellensteuerthemen als auch bei der schnellen Erstellung von neuen Verträgen helfen – natürlich immer unter konsequenter Einhaltung datenschutzrechtlicher Bestimmungen und Gewährleistung der Vertraulichkeit.

Bei Unternehmenskäufen und -verkäufen findet im Rahmen einer Tax Due Diligence bisher meistens ein manueller Review des Datenraums statt. Doch die Daten hier sind oft unvollständig und unsortiert, weshalb sich der Review oft zeitintensiv gestaltet. Hinzu kommt, dass die meisten Mandanten immer gebührensensibler werden. Hier bietet sich ein wunderbarer Anwendungsfall für KI und generative Sprachmodelle. Denn diese sind in der Lage, eine Vielzahl von Daten zu prüfen, Nachfragen zu stellen und am Ende einen Bericht zu den Rückschlüssen auf potenzielle Risiken zu erstellen.

Und auch auf Beraterseite bieten sich weitere Anwendungsfälle von KI: zum Beispiel im Rahmen der Erstellung einer Steuererklärung. Hier kann eine KI eingesetzt werden, um automatisierte Anforderungslisten zu erstellen, erhaltene Daten mit den Vorjahren abzugleichen und Abweichungen hervorzuheben. Ist die Steuererklärung erstellt, kann eine KI auf der Basis ein Musteranschreiben erstellen und darin z. B. die Abweichungen zwischen der Handelsbilanz und der Steuerbilanz darstellen.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Herausforderungen werden vor allem bei den Skills der Mitarbeiter gesehen. Denn gerade die erfahrenen Mitarbeiter müssen bei Neuerungen mehr an die Hand genommen werden als junge Neueinsteiger. In diesem Zusammenhang ist vor allem das Thema Change Management sehr wichtig, damit KI Projekte überhaupt erfolgreich sein können.

Ein weiterer Schwerpunkt ist das Thema Datenschutz, das gerade in Europa mit seinen strengen Regularien besonders im Fokus steht. Hier wird eine aufwendige Anonymisierung von sensiblen Daten notwendig sein, damit keine vertraulichen Daten offengelegt werden. Lediglich den Unternehmensnamen zu schwärzen, reicht hier wohl kaum aus, wenn die restliche Beschreibung weiterhin sehr detailliert ausfällt.

Dass Sprachmodelle auch halluzinieren und auf nichtexistierende Quellen verweisen können, hat man bereits in der Vergangenheit gehört. Dies kann dem Vertrauen in eine KI schaden und eine gewisse Skepsis schaffen. Es zeigt aber auch, dass man sich mit dem Thema der Haftung intensiv beschäftigen muss. Denn wer haftet eigentlich, wenn man sich auf manipulierte Informationen verlässt?

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Natürlich muss man für die Einführung von KI-Tools auch Geld investieren. Selbst dann, wenn sie aktuell noch nicht zwingend den erkennbaren Mehrwert mit sich bringen. Hierbei werden kleinere Mandanten mit geringeren Budgets wohl keine eigene KI basteln. Demgegenüber werden größere Mandanten eher tätig, da diese über entsprechende Kapazitäten verfügen. Dies ist eine Vermutung, die sich im aktuellen Markt auch klar beobachten lässt. Für den Mittelstand heißt es derzeit noch eher abwarten. Dies ist keine

schlechte Idee, da die aktuelle Entwicklung sehr rasant ist und sich steuerliche KI oft noch in einer Beta-Phase der Entwicklung befindet.



Was den Nachwuchs in den Steuerabteilungen angeht, wird der Markt für Leute mit technologischen als auch steuerlichen Fähigkeiten sehr umkämpft sein. Daher braucht es für KI-Themen in Zukunft ein „Learning on the Job“. Hier kann auch der Berater in Kooperation mit Softwareanbietern unterstützend aktiv werden und bspw. Schulungen anbieten.

3.5 Anwendungen im Bereich Zoll

Allgemeine Informationen

Zölle bezeichnen im Allgemeinen Abgaben, die bei der Überführung von Waren über Zollgrenzen erhoben werden. Unterschieden wird hierbei zwischen der Einfuhr von Waren in ein Zollgebiet (Einfuhrzölle) und der Ausfuhr von Waren (Ausfuhrzölle). Bei der Überschreitung von Zollgrenzen beim Warenimport wird im Rahmen einer Zollprüfung untersucht, ob Einfuhrabgaben anfallen und in welcher Höhe diese abgeführt werden müssen. Schwerpunkte dieser Tätigkeiten liegen bei der Einreihung oder Tarifierung von Waren. Dabei werden alle Waren mit einer eindeutigen Warennummer versehen, die nachfolgend der Berechnung des Zollwerts zugrunde liegen.

Die Korrektheit der Zollanmeldung wird seitens der Zollverwaltung überprüft. Inkorrekte Anmeldungen müssen umgehend korrigiert werden und können zu erheblichen Zollnachforderungen führen. In schwerwiegenden Fällen sogar zur Einleitung eines Strafverfahrens. Um eine korrekte zollrechtliche Tarifierung von Waren sicherzustellen, bestimmen viele Unternehmen daher häufig einen Zollbeauftragten. Dieser beschäftigt sich in der Regel mit allen zollbezogenen Themen.

Die Kosten für den Zoll müssen bei der Gestaltung von Verkaufspreisen berücksichtigt werden. Dazu müssen sämtliche zollrelevanten Daten digitalisiert werden, um eine maschinelle Auswertung zu ermöglichen. Oftmals sind im Rahmen der zollrechtlichen Behandlung auch andere steuerrechtlich relevante Aufgaben sowie Logistik- und Versandaufgaben und Tätigkeiten zur Einhaltung der innerbetrieblichen Exportkontrolle relevant.

Experteninterview

mit Alexandra Klein und Holger Bauer

FRAGE 1: ZOLL- UND EXPORTKONTROLLFUNKTION DER ZUKUNFT

Was konkret seht ihr, wenn ihr euch die Zoll- und Exportkontrollfunktion der Zukunft im Unternehmen vorstellt? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Für die Zoll- und Exportkontrollfunktion der Zukunft wird es entscheidend sein, dass sie die neuen technologischen Möglichkeiten zur Entscheidungsfindung einsetzt und weniger zur Erfüllung von Compliance-Aufgaben. Denn nur so kann sie auch als Funktion wahrgenommen werden, die durch eine schnelle Entscheidungsfindung und die effektive Nutzung von Daten einen spürbaren Mehrwert für das Management liefert. Um dies zu ermöglichen, werden neben dem Steuersachverständnis künftig auch andere Skills benötigt. Hierzu zählen eine gewisse IT-Affinität und Marktkenntnisse zu vorhandenen Systemen. So wird sich insgesamt das Berufsbild in Richtung eines Technologieberaters entwickeln, der sich auch mit steuerlichen Grundlagen auskennt. Reines Excel-Know-how wird hier nicht mehr ausreichend sein. Nur mit dieser Entwicklung wird die Steuerfunktion in Zukunft sichtbarer im Unternehmen – gerade auch im Vergleich zu anderen Bereichen.

Dabei darf auch das Thema Data Governance nicht vergessen werden. Nur wenn der entsprechende Input für die Systeme in guter Qualität vorhanden ist und die Informationen belastbar sind, kann eine KI vernünftig eingesetzt werden. Dies ist in der Zoll- und Exportkontrollfunktion umso wichtiger, da Informationen aus verschiedenen Quellen des Unternehmens kommen und auch teilweise bei Dienstleistern außerhalb des Unternehmens liegen.

Darüber hinaus erwarte ich in der Zukunft eine stärkere Fokussierung auf das Thema Datenanalyse bei der Zoll- und Exportkontrollfunktion. Denn gerade in diesen Bereichen existieren heutzutage zum Teil noch immer mangelnde Analysemöglichkeiten bei großen Datenmengen. Dies ist eine Entwicklung, die auch

von der Finanzverwaltung zu erwarten ist. Diese könnte mit fortgeschrittenen Analysemöglichkeiten den Prüfungsfokus besser setzen, was für Unternehmen ein entsprechendes Risiko darstellen kann.

Das Anforderungsprofil an die Zollexperten wird sich mehr in Richtung des technischen Know-hows entwickeln. In der Folge wird es in Zukunft keinen zusätzlichen IT-Experten mehr brauchen, um notwendige Reports aus den verschiedenen Systemen zu extrahieren und eine Analyse der Daten durchzuführen. Doch dafür muss sich auch die Lehre und Ausbildungsstrategie entsprechend anpassen. Aktuell lassen sich in der Ausbildung oder im Studium der Zollverwaltung wenig bis gar keine Bezüge zu Automatisierung und KI finden. Hinzu kommt, dass im Zollbereich auch viele Quereinsteiger unterwegs sind. Heißt konkret: Es bieten sich auch jede Menge Chancen und gute Perspektiven für ein durchaus interessantes Arbeitsumfeld.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

Hier muss eine Zollorganisation in zwei Bereiche unterscheiden werden: Im nicht-operativen Zollbereich, wie der steuerrechtlichen/zollrechtlichen Beratung, kam KI in den vergangenen Jahren eher weniger zum Einsatz. Im Bereich der Zolltarifizierung hingegen hat sich seit 2017 einiges getan. Hier haben vor allem SAP-Partnerunternehmen entsprechende Lösungen vorgestellt, um Prozesse bei der Auslesung von Informationen zur Tarifizierung mit Hilfe von KI weiter zu verbessern. Auch Prüflogiken wurden etabliert, mit denen sich Abweichungen in Zollanmeldungen effektiv identifizieren lassen. So werden bspw. Importzollanmeldungen automatisch mit Zollanmeldungen der vergangenen Monate abgeglichen und dem Sachbearbeiter anhand von prozentualen Abweichungen präsentiert. Dies stellt einen Teilbereich eines internen Kontrollsystems (IKS) dar, bevor die Zollanmeldung abgegeben wird.

Im Bereich der Tarifizierung ist auch die WTS ClassificationEngine zu nennen. Bei diesem Tool werden sowohl gesetzliche Grundlagen aus dem Zolltarif und der Exportkontrollgüterlisten (Stichwort: Russland Sanktionen!) als auch mandantenspezifische Informationen (Attribute) genutzt, um bei der Tarifizierung zu unterstützen. Die WTS ClassificationEngine verbindet dabei nicht nur Machine Learning mit einem individuell einstellbaren Regelmechanismus. Das Tool ist auch fester Bestandteil unserer Arbeitsabläufe im digitalen Zoll Business Partnering, um Mandanten in verschiedenen Industriesektoren effektiv betreuen zu können.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Nach den zollrechtlichen Bestimmungen ist es wichtig, dass die Ware bei der Anmeldung korrekt angesprochen und nicht in Abkürzungen aufgeführt wird. Hier zeigt sich vor allem für generative Sprachmodelle ein großes Potenzial, wenn diese im internationalen Kontext genutzt und global zum Einsatz kommen. So lassen sich aus dem Kontext heraus korrekte Übersetzungen liefern und einsetzen. Dabei ist es natürlich wichtig, dass die Qualität sichergestellt wird. Die Aufgabe dieser Qualitätssicherung kann wiederum die Zollabteilung übernehmen.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Das größte Potenzial zur Nutzung von KI liegt sicher im operativem Zollbereich. Hierzu zählen zum Beispiel die automatisierte Erstellung von Zollanmeldungen und das damit verbundene System zur Prüfung der Zollanmeldungen (Stichwort: IKS). Auch ist ein weiterer Ausbau bei der Zolltarifizierung möglich und verbessert hier die Qualität. Ein lernendes System kann sehr wertvoll sein, das eine Plausibilisierung durchführt und prüft, welche Produkte bereits an einen Kunden exportiert wurden oder bei welchen Produkten es Abweichungen zum bisherigen Verhalten gibt. Vor allem dann, wenn es um die Einhaltung von Sanktionsvorgaben geht. Es kann aber auch wertvoll für die Enthftung bei etwaigen Zollverstößen sein. Damit wird die Compliance für Unternehmen maßgeblich verbessert.

Doch auch bei der zollrechtlichen Beratung ergeben sich durch die generativen Sprachmodelle neue Möglichkeiten. Da bei der Erstellung von Gutachten meist verschiedene Quellen zum Einsatz kommen, kann die KI sowohl bei der Recherche als auch bei der Formulierung von Texten helfen. So müssen zum Beispiel bei der Frage des zollrechtlichen Ausführers regelmäßig Gesetze, Kommentierungen, Dienstvorschriften sowie Schreiben und Infos der Behörden durchsucht werden. Hier kann ein generatives Sprachmodell für eine signifikante Zeitersparnis sorgen.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Das Thema Data Governance ist eine der zentralen Herausforderungen beim Thema KI, da die Systeme immer nur so gut sind wie deren Inputdaten. In diesem Zusammenhang werden Schnittstellen und Datenqualität zunehmend wichtiger. Denn eine KI-Lösung kann immer nur dann helfen, wenn auch die zugrundeliegenden Stammdaten gepflegt werden. Dafür müssen Zollexperten allerdings ihre Denkweise verändern und Datenwächter werden. Denn nur wenn Unternehmen lernbegierig sind und aus der transaktionalen Sicht herauskommen, können sie den Fokus auch auf spezifische Fälle legen, die in der Vergangenheit oft zu Problemen geführt haben.

Doch auch unter Berücksichtigung des enormen Potenzials von KI sollte die Geschäftsleitung nicht denken, dass in Zukunft keine Mitarbeiter mehr mit fachlichen Kenntnissen gebraucht werden. Selbst wenn ein Großteil der Arbeit durch die KI abgedeckt wird, braucht es dahinter noch immer einen Menschen zur Plausibilisierung mit fachlichem Know-how.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Um für die Zukunft gerüstet zu sein, müssen Zoll- und Exportkontrollabteilungen vor allem einen Fokus auf die Organisationsthematik legen. In einem funktionierenden Team braucht es daher eine Kombination aus zollfachlichem Know-how, prozessuellem Know-how und technischem Know-how. Diesen Skill-Mix gilt es, stets in der Balance zu halten. Externe Berater haben oft einen guten Marktüberblick für technische Lösungen, der für Unternehmen zum Teil nur schwer zu erlangen ist. Dadurch sind sie in der Lage, als technologischer Sparringspartner aufzutreten, der strategisch unterstützt und gezielt im operativen Bereich zum Einsatz kommt.

3.7 Anwendungen im Bereich Verrechnungspreise

Allgemeine Informationen

Der Begriff Verrechnungspreis bezeichnet die monetäre Bewertung von Wirtschaftsgütern, die innerhalb einer Organisation zwischen selbstständigen Bereichen in unterschiedlichen Ländern ausgetauscht werden. Hierzu zählen sowohl materielle Güter wie z. B. Warenlieferungen als auch immaterielle Güter wie bspw. Lizenzen oder Nutzungsrechte. In der Regel können diese Leistungen nicht auf dem freien Markt erworben werden, weshalb zur korrekten Preisbestimmung verschiedene Verfahren zur Bewertung vergleichbarer Leistungen angewendet werden müssen. Von dieser Thematik sind besondere Konzerne mit organisatorisch getrennten Unternehmensbereichen bzw. unterschiedlichen Unternehmen betroffen. Ziel ist es, dass die Gesellschaften und Betriebsstätten in verschiedenen Ländern ihren jeweiligen Gewinn gesetzeskonform berechnen und versteuern.

Aufgrund gesetzlicher Regelungen sind Unternehmen dazu verpflichtet, eine ausführliche Dokumentation der angewendeten Methodik nachzuweisen, mit der sie ihre Verrechnungspreise ermitteln. Verstöße gegen diese Regelungen können zu hohen steuerlichen Nachzahlungen und Strafen führen. Für international agierende Konzerne können die Ausnutzung und Optimierung von Verrechnungspreisen einen hohen Einfluss auf die Gewinnermittlung und die Besteuerung haben. Gleiches gilt für das Konzernergebnis, das ebenfalls maßgeblich beeinflusst wird.

Experteninterview mit Andreas Riedl

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret siehst du, wenn du dir die Steuerfunktion der Zukunft im Unternehmen vorstellst? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

In der Zukunft wird es nicht mehr nur ein Skillset in der Steuerabteilung geben, da gerade im TP-Bereich rein steuerliches Fachwissen längst nicht mehr ausreicht. Vielmehr braucht es auch ökonomisches Wissen, das passend zu den betriebswirtschaftlichen Hintergründen gefragt ist. Darüber hinaus hat man in den letzten Jahren gesehen, dass Technologie- und Prozessverständnis eine immer größere Rolle spielen.

Die Steuerfunktion wird in Zukunft mit wachsenden Compliance-Anforderungen konfrontiert werden. Um diesen Herausforderungen gerecht zu werden, spielen Technologien zwar eine wichtige Rolle – allerdings keine Hauptrolle. Die zentrale Herausforderung und zugleich Kernfunktion der Steuerfunktion der Zukunft wird im Management des Qualifikationsmixes der Mitarbeiter liegen. Steuerabteilungen, die ausschließlich Steuerrechtler anstellen, werden in Zukunft nicht mehr funktionieren.

Ich gehe davon aus, dass die Arbeit durch technologische Innovationen in den kommenden Jahren eher erweitert wird anstatt reduziert zu werden. Begründen lässt sich diese Annahme durch die immer höheren Compliance-Erfordernisse und den Mangel an Technologieexpertise, die für einen großen Übersetzungsbedarf an dieser Schnittstelle sorgen werden. Bei Verrechnungspreisen wird es zukünftig einfacher Sollbruchstellen zu finden. Technologieaffine Betriebsprüfer werden schneller erkennen, sollte die Realität nicht mit der Dokumentation übereinstimmen. Um dem entgegenzuwirken, müssen Steuerabteilungen noch integrativer mit dem Business zusammenarbeiten.

FRAGE 2: BLICK ZURÜCK

Wie schätzt du den Mehrwert von KI mit Blick auf die vergangenen Jahre in deinem Bereich ein – also die Zeit vor der Ära der generativen KI?

Was KI-Anwendungen betrifft, kann man diese für die vergangenen Jahre in zwei Bereiche unterscheiden:

Zum einen wurden bereits für größere Themenbereiche KI Modelle selbst entwickelt, um vordefinierte Aufgaben zu bewältigen. Doch viele dieser Anwendungen befinden sich noch im Versuchsstatus und die Häufigkeit der Nutzung fiel bisher vergleichsweise gering aus. Auf der anderen Seite hat die KI bereits Einzug in die Arbeitswelt gehalten. So kamen in der breiten Masse verschiedenste, kleine Tools zum Einsatz und haben die Arbeit extrem erleichtert. Immer häufiger werden bspw. KI-gestützte Übersetzungsprogramme verwendet, anstatt Übersetzungsbüros zu beauftragen. Hier hat die Disruption also bereits stattgefunden.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt du aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben deiner Ansicht nach generative Sprachmodelle?

Auch mit Blick auf die Gegenwart kann man zwei Ebenen unterscheiden: So wird KI in verschiedenen Anwendungen weiter in die Arbeit einfließen, ohne dass es der Nutzer groß bemerkt. Schon bald wird es zum normalen Alltag gehören, dass Texte von einer KI Textkorrektur gelesen werden. Oder E-Mails formuliert und Excel-Formeln generiert. Auf der anderen Ebene wird KI die fachliche Arbeit inhaltlich prägen. So ist es durchaus möglich, dass im Verrechnungspreisbereich zukünftig hypothetische Fremdvergleiche mit vordefinierten Parametern von einer KI erstellt werden. Alternativ kann sie aber auch bei der Verteidigung in Betriebsprüfungen zur Anwendung kommen.

Auch Prüfer werden auf die Möglichkeiten von KI zurückgreifen um Verrechnungspreise zu kontrollieren und plausibilisieren. Ob diese Möglichkeiten insgesamt dabei helfen können, steuerliche Konflikte zu lösen oder doch eher für eine höhere Komplexität sorgen, bleibt noch abzuwarten. In Summe lässt sich jedenfalls jetzt schon festhalten, dass sich die Anwendungsmöglichkeiten von KI durch die aktuelle Entwicklung der generativen Sprachmodelle enorm gesteigert haben.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in deinem Bereich siehst du aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Es ist davon auszugehen, dass KI überall im TP-Lebenszyklus eine relevante Rolle spielen wird. Was dazu führen wird, dass man in Zukunft nicht mehr von separaten Anwendungsfällen spricht. Doch wirft man einen Blick auf konkrete Bereiche, ist durch die generativen Sprachmodelle vor allem die Anwendung bei der Dokumentation von Verrechnungspreisen interessanter geworden. Durch gezielte Prompts können hier Texte für Industrieanalysen und Unternehmensbeschreibungen relativ schnell und einfach erstellt werden.

Auch die Erstellung ganzer Local Files und Master Files auf Knopfdruck erscheint in Zukunft nicht mehr unrealistisch. Daneben kann die KI aber auch bei Betriebsprüfungen zum Einsatz kommen, um Argumente zu finden, Korrelationen aufzudecken und eine konzernweite Konsistenz bei der Beantwortung von Anfragen sicherzustellen. Nicht zu vergessen die Erstellung einer Fremdvergleichsanalyse als klassischen Anwendungsfall, der mit Hilfe von KI weiterhin relevant bleiben wird.

Doch neben der KI haben sich auch andere Technologielösungen wie z. B. die sogenannten Low-Code- bzw. No-Code-Lösungen stetig weiterentwickelt. So ist es mittlerweile deutlich schwerer geworden zu durchschauen, in welchen Anwendungen tatsächlich KI verwendet wird. Bereits jetzt haben viele Konzerne

die Übersicht über lizenzierte Tools und Lösungen verloren. Ein Zustand, der sich durch die neuen KI-Lösungen nur weiter verschärfen wird. Aufgrund dessen werden Berater in Zukunft auch immer häufiger eine Wegweiserfunktion übernehmen.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in deinem Bereich voll ausschöpfen zu können?

Die größte Herausforderung ist die Zeit. Denn die meisten Mandanten und Berater bedienen Compliance-Anforderungen aus dem Hier und Jetzt heraus, wodurch sie vollumfänglich ausgelastet sind. Hinzukommend wird es immer schwieriger, qualifiziertes Personal auf dem Arbeitsmarkt zu finden, um diese Themen weiter voranzutreiben. Datenschutz, Erklärbarkeit der Ergebnisse und Fähigkeiten der Mitarbeiter – das sind nur einige der Herausforderungen, die deutlich machen, warum das Thema integrative Technologienutzung eine strategische Priorität im Unternehmen und in der Steuerabteilung bekommen muss – auch und vor allem wegen der Einbeziehung von KI.

Doch abseits der großen DAX-Unternehmen ist man in einer Eieruhr gefangen. Zwar liefern neue Technologien viele Möglichkeiten, um die Arbeit effizienter und effektiver zu gestalten. Allerdings braucht es Zeit und entsprechende Kapazitäten, um diese zu implementieren. Das Bild hat sich im Vergleich zur Digitalisierung und Automatisierung ohne KI nicht geändert. Wenn man nie anfängt, kommt man auch nicht voran. Und wenn man nicht genug Fahrt aufnimmt, entwickeln sich Technologien schneller als man selbst überhaupt in die Umsetzung kommt.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Hier sind zwei Punkte entscheidend: Zum einen müssen Steuerabteilungen einen größeren Fokus auf ein diverseres Skillset ihrer Mitarbeiter legen und nicht nur z. B. Fachanwälte für Steuerrecht einstellen. Eine Tatsache, die auch abseits der KI-Entwicklung für Technologie an sich gültig ist. Personen mit Fachwissen aus dem Technologiebereich können als Katalysatoren innerhalb einer Abteilung fungieren, sofern sie von einem aktiven Change Management begleitet werden. Wenn hier Personen mit einem bestimmten Skillset kurz- oder mittelfristig fehlen, können wir als Berater aushelfen und Lücken mit den Unternehmen schließen, um den Wandel aktiv zu unterstützen.

Auf der anderen Seite ist es enorm wichtig, von Erfolgsgeschichten zu lernen und diese zu kopieren. So können sich Unternehmen mit anderen Betrieben aus der Region bzw. einer anderen Branche zusammenschließen, um sich über die Steuerabteilung hinaus untereinander auszutauschen. Was letztlich zu einer gegenseitigen Befruchtung führen und bei der individuellen Lösungsfindung unterstützen kann. Dabei ist es Aufgabe der Berater, die Unternehmen hierfür zu sensibilisieren und als Kontaktvermittler zu anderen Unternehmen aufzutreten. Wer hier am Markt starke Referenzen aufweisen kann, wird in diesem Bereich führend sein.

3.9 Anwendungen im Bereich Tax CMS / steuerliches Risikomanagement

Allgemeine Informationen

Kernelement eines steuerlichen Risikomanagements bzw. Tax CMS ist die Etablierung eines Prozesses zur systematischen Identifikation und Bewertung von steuerlichen Risiken. Hierzu zählt auch die Implementierung entsprechender Maßnahmen zur Reduzierung der Risiken.

Sofern die steuerrelevanten Prozesse nicht ordnungsgemäß organisiert sind, kann es dazu kommen, dass die steuerlichen Pflichten eines Unternehmens nicht ordnungsgemäß erfüllt werden. Dies kann strafrechtliche Sanktionen nach sich ziehen. Zudem können sich persönliche Haftungsrisiken z. B. für Geschäftsführer, Vorstände oder Leiter von Steuerabteilung ergeben. Insbesondere die sehr komplexen und vielgestaltigen, regulatorischen Vorgaben, sich dynamisch verändernde Geschäftsmodelle sowie eine fortschreitende Digitalisierung sämtlicher Prozesse im Unternehmen stellen das steuerliche Risikomanagement vor große Herausforderungen.

Dazu kommt: Steuern betreffen nahezu jeden Prozess im Unternehmen, denn überall müssen steuerrelevante Entscheidungen gefällt werden. Insoweit beschränkt sich das steuerliche Risikomanagement bzw. Tax CMS nicht allein auf die Prozesse innerhalb der Steuerabteilung. Vielmehr muss es sich auf alle Prozesse bzw. Prozessschritte erstrecken, die eine steuerliche Relevanz haben können. Ein umfassend aufgebautes steuerliches Risikomanagement bzw. Tax CMS reduziert das Risiko von Fehlern und hat – sollten dann doch einmal Probleme auftreten – idealerweise eine enthaftende Wirkung. Damit schützt es die Verantwortlichen sowie die Mitarbeiter und am Ende das Unternehmen.

Experteninterview

mit Christiane Belz und Ellen Birkemeyer

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret seht ihr, wenn ihr euch die Steuerfunktion der Zukunft im Unternehmen vorstellt? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Die technologischen Entwicklungen werden auch die Stellung der Steuerfunktion im Unternehmen und ihre Arbeit verändern. In der Vergangenheit war die Steuerabteilung sehr stark retrospektiv tätig und hat bereits abgeschlossene Sachverhalte analysiert und z. B. für Zwecke der Steuererklärung verarbeitet. Das war oftmals eine sehr manuelle Arbeit. Diese Arbeitsweise funktioniert jedoch zunehmend dann nicht mehr, wenn sich der Rest des Unternehmens digitalisiert und die Prozesse immer automatisierter ablaufen.

Um sicherzustellen, dass die automatisierten Prozesse nahtlos und effizient sowie aus steuerlicher Sicht zutreffend ablaufen, muss sich die Steuerabteilung mit ihren (steuerlichen) Vorgaben in den Prozessen „verankern“, denn oftmals werden die steuerrelevanten Entscheidungen bereits am Anfang des Prozesses getroffen. Wenn dort etwas falsch läuft, weil z. B. ein Sachverhalt unzutreffend steuerlich gewürdigt wird, ist das Risiko hoch, dass sich dies am Ende in einer falschen Deklaration niederschlägt. Insoweit ist die Steuerabteilung bereits heute viel intensiver in die operativen oder administrativen Abläufe des Unternehmens eingebunden als in der Vergangenheit. Diese Entwicklung wird sich weiter fortsetzen.

Der sich verändernde Fokus hat natürlich auch Einfluss auf das Skillset der Mitarbeiter der Steuerabteilung: Prozess- und IT-Know-how werden sehr viel wichtiger und neben Steuerexperten wird die Steuerabteilung in der Zukunft vor allem aus IT-affinen Mitarbeitern bestehen, die z. B. über eine besondere Expertise im Bereich der Datenanalysen oder über eine umfassende SAP-Kenntnis verfügen.

Zudem wird die Arbeit innerhalb der Steuerabteilung zunehmend weniger nach Spezialisierungen als vielmehr entlang der unternehmensinternen Prozesse organisiert werden. Dies schafft dann auch die Grundlage für einen sinnvollen Einsatz von KI in den steuerrelevanten Prozessen.

Insoweit werden sich in der Zukunft viele Anwendungsmöglichkeiten von KI in allen Steuerbereichen ergeben, nicht nur um effizienter zu werden, sondern auch um die Compliance stärker zu unterstützen. Letzteres kann z. B. durch „intelligente“ Prüfroutinen erreicht werden, mittels derer Auffälligkeiten identifiziert werden.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

Bei der Frage ist es wichtig einzuordnen, wie die KI in der Vergangenheit gesehen wurde. Im klassischen Tax CMS Bereich konnte man das Thema KI noch nicht so ganz einordnen. Insbesondere wurde KI bisher auch nicht in den übrigen Prozessen eingesetzt und sie wurde daher auch in unserem Bereich so gut wie nicht genutzt.

Was man in den vergangenen Jahren allerdings gesehen hat ist, dass sich viel in Richtung Automatisierung von Prozessen getan hat. So versucht man zum Beispiel über Process Mining die Ist-Prozesse zu analysieren und darüber auch steuerliche Risiken zu identifizieren anstatt diese manuell aufzunehmen.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Das Potenzial von KI im Bereich Tax CMS ist enorm. Zum einen kann KI uns im Rahmen der Analyse von Schwachstellen helfen; zum anderen bei der Analyse von fachlichen Themen z. B. zur Unterstützung von Entscheidungsmodellen. Insoweit kann KI dabei unterstützen, die Prozesse robuster und sicherer zu machen.

Darüber hinaus wird die KI auch einen Einfluss auf Betriebsprüfungen haben, da auch die Finanzverwaltung an neuen Methoden zur Analyse von Prozessen arbeitet. Dementsprechend kann die KI maßgeblich dazu beitragen, Prüfungsschwerpunkte zu identifizieren oder zu validieren. Dabei soll herausgefunden werden, inwiefern Feststellungen abgeholfen wurden, die in einer Vor-BP getroffen worden sind. Zudem wird bereits heute eine Vielzahl von Daten abgefragt, die bisher kaum ausgewertet werden. Dies wird sich mit Einsatz von KI sicherlich grundlegend ändern und darauf müssen sich die Unternehmen vorbereiten.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Grundlage eines Tax CMS ist die Identifikation der steuerlichen Risiken. Um diese zu ermitteln, müssen zunächst die Prozesse analysiert werden. Im Rahmen dieser Analyse wird man zukünftig verstärkt KI z.B. im Rahmen von Process Mining einsetzen. Zudem kann KI im Rahmen der Überwachung unterstützen. Denn, wenn die entsprechenden Risiko-mitigierenden Maßnahmen einmal definiert sind, müssen diese hinsichtlich ihrer Durchführung überwacht werden. D.h. es sind sinnvolle Kontrollen zu implementieren,

die eine effiziente Überwachung ermöglichen. Auch dabei kann KI unterstützen, um z.B. Anomalien zu identifizieren. Diesen Anomalien kann dann gezielt nachgegangen werden.

Zudem kann KI helfen, andere Abteilungen, die steuerrelevante Sachverhalte beurteilen, jedoch nicht über das entsprechende steuerliche Know-how verfügen, im Rahmen der Entscheidungsfindung zu unterstützen. Hier sehen wir ein großes Potential im Bereich der transaktionalen Steuern, wie der Lohnsteuer oder der Umsatzsteuer. Denkbar wäre z. B. der gezielte Einsatz von Chat Bots für die lohnsteuerliche Beurteilung von Sachzuwendungen oder die Einordnung von Verträgen im Zusammenhang mit Quellensteuern. Im Vergleich zu starren Entscheidungsbäumen oder gar Richtlinien kann KI hier zu einer wesentlichen Prozessverbesserung und damit insbesondere auch zu einer höheren Akzeptanz führen.

Ein weiterer möglicher Anwendungsfall für den Einsatz von KI im Rahmen von Tax CMS ist das „Legal monitoring“. Dabei ist ein System aufzusetzen, in dessen Rahmen rechtliche Änderungen und deren Einfluss auf die Prozesse analysiert werden. KI kann sowohl im Rahmen der Analyse als auch bei dem Erarbeiten von konkreten Maßnahmen unterstützen.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Neben dem großen Thema Datenschutz ergeben sich Herausforderungen vor allem im Zusammenhang mit Haftungsfragen. Wenn zum Beispiel eine steuerliche Frage durch einen KI gestützten Entscheidungsbaum falsch beantwortet und damit steuerlich unzutreffend beurteilt wird: Wer trägt dann dafür die Verantwortung? Kann man sich enthaften, indem man sich auf die KI beruft?

Aber auch das Thema der Prozesstransparenz ist ein wichtiges Thema. Nur wenn man seine Prozesse kennt, kann man beurteilen wo z. B. KI unterstützte Entscheidungsbäume sinnvoll einsetzbar sind. Muss die Transparenz zunächst hergestellt werden, stellt dies eine hohe Hürde dar.

Ein weiterer entscheidender Aspekt ist die Dokumentation. Wenn eine KI genutzt wird, muss auch dokumentiert werden, wie es zu einer Entscheidung gekommen ist. Das muss im Zweifelsfall auch noch Jahre später im Rahmen einer Betriebsprüfung erläutert werden können.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Wie bereits eingangs diskutiert, wird sich die Rolle einer Steuerabteilung weiter verändern. Dies hat Einfluss auf die Anforderungen, die an die Mitarbeiter in einer Steuerabteilung gestellt werden. Neben der steuerfachlichen Kompetenz, müssen diese vor allem eine grundsätzliche Prozess- und Technologieaffinität mitbringen.

Als Berater können wir die Steuerabteilung bzw. ihre Mitarbeiter im Rahmen des Veränderungsprozesses unterstützen, indem wir z. B. KI-unterstützte Anwendungsfälle entwickeln. Zudem könnte man darüber nachdenken, bestimmte Prozesse an Berater auszulagern. Wenn z. B. mehrere Unternehmen ihre Verträge im Hinblick auf Quellensteuerrelevanz von einem Berater mittels einer speziellen - KI-unterstützten - Software untersuchen lassen, könnte dies effizienter sein, als wenn eine solche Prüfung durch jedes Unternehmen selbst erfolgt.

Insoweit würde die KI mit größeren Datenmengen „gefüttert“ und könnte dadurch schneller trainiert werden. Als Berater hat man aber auch die Aufgabe, über Risiken und Nebenwirkungen eines KI-Einsatzes aufzuklären und aufzuzeigen, wo eine KI möglicherweise (noch) nicht sinnvoll eingesetzt werden kann.

Zudem kann man als Berater dabei unterstützen, Regeln für den Einsatz von KI zu definieren z.B. im Zusammenhang mit den vorzunehmenden Qualitätssicherungen.

3.10 Anwendungen im Bereich International Tax

Allgemeine Informationen

Der Bereich Internationale Projektberatung beschäftigt sich u. a. mit der steuerlichen Betreuung von Mandanten im Fall von grenzüberschreitenden operativen Projekten. Mandanten, die an Ausschreibungen im internationalen Projektgeschäft teilnehmen wie z. B. bei Bau- und Montageprojekten, benötigen vor Abgabe eines Angebots eine vorläufige Abschätzung der steuerlichen Konsequenzen. Denn diese müssen in der Ausschreibung entsprechend berücksichtigt werden.

Die steuerlichen Rahmenbedingungen im Projektstaat und die zunehmende Komplexität im internationalen Umfeld stellen in grenzüberschreitenden Projekten besondere Herausforderungen an die steuerliche Betreuung in den jeweiligen Projektphasen. Die steuerliche Betrachtung und Risikobewertung reicht hierbei über verschiedene Steuerarten hinweg und betrifft sowohl Fragen der Ertragssteuern wie z.B. Betriebsstätten oder Quellensteuern und die Umsatzsteuer. Darüber hinaus schließt es die Besteuerung der Einkommen der eingesetzten Mitarbeiter und deren Sozialversicherung mit ein.

Bereits in der Angebotsphase eines Projekts und den vertraglichen Verhandlungen müssen die speziellen steuerlichen Rahmenbedingungen berücksichtigt werden, die wiederum Einfluss auf die Projektgestaltung und -realisierung haben. Im Falle eines Zuschlags bei einer Ausschreibung müssen anschließend die steuerlichen Konsequenzen des Projektes vor Ort nachvollzogen und geprüft werden.

Typische anfallende Tätigkeiten im Rahmen der Projektdurchführung sind die Unterstützung bei der Registrierung von ausländischen Betriebsstätten, die Erstellung von Steuererklärungen und die Unterstützung bei Betriebsprüfungen. Besondere Herausforderungen im Rahmen dieser Tätigkeiten sind lokale Gesetzesbestimmungen sowie sprachliche und kulturelle Barrieren.

Experteninterview

mit Ulrike Schellert und Martin Loibl

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret seht ihr, wenn ihr euch die Steuerfunktion der Zukunft im Unternehmen vorstellt? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Die Beantwortung der Frage ist aus unserer Sicht abhängig von Philosophie und Größe der Steuerabteilung. Wieviel möchte ich selber machen, wieviel möchte ich auslagern, z. B. über ein Business Partnering, und welche Personen brauche ich für diese Aufteilung? Letzteres ist ein Mix aus Steuerspezialisten mit fachspezifischem Detailwissen und steuerlichen Generalisten mit zusätzlichem Wissen über die Unternehmensprozesse.

Bezogen auf die Wahrnehmung des Berufsbildes kommt schnell die Frage auf: Braucht es die Steuerabteilung in Zukunft überhaupt noch? Oder macht am Ende alles die KI? Hier muss man aus meiner Sicht darauf achten, dass die Steuerabteilungen nicht abgewertet werden. Denn eine KI ist immer nur so schlau, wie die Daten, mit der sie gefüttert wird. Für die Interpretation der Daten und den Review der Ergebnisse ist der Mensch nach wie vor notwendig.

Doch was die Entwicklung in der Zukunft angeht, denken wir noch immer zu klein. Diese wird noch viel rasanter voranschreiten. Trotzdem bleibt die Befürchtung, dass wir uns heute über tolle Ideen unterhalten

und dann aber wieder 5 Jahre ins Land gehen, ohne dass überhaupt etwas passiert. Umso wichtiger ist es daher, die Anwendungsfälle jetzt schnell in ein nutzbares Produkt zu überführen, das in die Arbeit integriert werden kann. Denn daraus können wieder neue Ideen entstehen und es lassen sich die Leute mitnehmen, die aktuell mit KI noch eher weniger zu tun haben.

FRAGE 2: BLICK ZURÜCK

Wie schätzt ihr den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

Aus meiner Sicht hat sich in den vergangenen Jahren noch nicht viel im Bereich International Tax getan. Zwar hat KI an sich Einzug in unser Leben erhalten wie z. B. mit Apps für Navigation. Teilweise sogar ohne unser Wissen. Allerdings ist hier mehr passiert als im Steuerbereich. Das Thema KI war in der Vergangenheit nicht jedem zugänglich. Auch wenn bereits vielen klar war, wohin die Reise in Zukunft einmal gehen könnte. So hatte man beim Blick auf die WTS Studie 2017 schon das Gefühl, dass eine Menge möglich ist. Doch die eigentliche Realisation hat sich dann als recht zäh herausgestellt.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt ihr aktuell das Potenzial für KI in eurem Bereich ein und welchen Impact haben eurer Ansicht nach generative Sprachmodelle?

Blickt man im Vergleich zur Vergangenheit nun auf die Gegenwart, dann herrscht eine gewisse Aufbruchsstimmung. Jetzt sind die KI-Lösungen auf einmal viel besser zugänglich und es kommen gewisse AHA-Momente auf, weil sich mit Hilfe der KI schon recht zügig ein tatsächlicher Mehrwert erzielen lässt. Ein Treiber dieser Entwicklung ist sicher die spielerische Herangehensweise. KI wird nicht mehr nur von Top-Experten beherrscht, sondern kann von jedem Einzelnen per App auf dem Smartphone genutzt werden.

Beim Potential der generativen Sprachmodelle muss man darüber hinaus zwischen der Steuerberatung und der Steuer-Compliance unterscheiden. Bei der Steuerberatung geht es weniger darum, mit Zahlen zu rechnen und diese in Formulare zu übertragen. Vielmehr geht es viel um Sprache, besonders bei der steuerlichen Würdigung von Sachverhalten. Genau deshalb wird sich in diesem Bereich noch vieles tun, da die Sprachmodelle vor allem im Umgang mit Texten ihre Stärken haben.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in eurem Bereich seht ihr aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Im Rahmen der steuerlichen Beratung haben wir in unserem Bereich etwa 5.000 Fälle im Jahr, die wir bearbeiten und dabei eine Menge Daten produzieren. Darunter sind viele repetitive Fälle bzw. Fälle mit einer großen Ähnlichkeit. Hier besteht ein perfekter Anwendungsfall für KI. Sofern man es schafft, das vorhandene Projekt-Know-how mit dem strukturierten Wissen aus steuerlichen Datenbanken zu kombinieren, um Antworten vorzubereiten. Denn aktuell findet gerade der Abgleich mit vergangenen Projekten häufig auf manueller Basis statt. Oder mit Hilfe einer simplen Stichwortsuche.

Ein anderer Anwendungsfall ist das Thema „Work from anywhere“, was aktuell alle Unternehmen aus allen Branchen betrifft. Ähnlich dem Beratungsgeschäft, hat man auch hier eine große Masse von Fällen mit ähnlichen Hintergründen, für die vordefinierte Steuerbereiche wie Körperschaftsteuer, Lohnsteuer und Sozialversicherung abgefragt werden müssen. Hinzu kommt, dass bei diesem Thema das Interesse eher bei den Mitarbeitenden liegt und die Budgets auf Unternehmensseite geringer ausfallen. Eine effektive Beantwortung dieser Fälle mit Unterstützung einer KI wird damit umso wichtiger.

Betrachtet man die Steuerabteilung im Unternehmen, lässt sich eines festhalten: Andere Bereiche kommen oft nur ungern mit Fragen auf die Steuerkollegen zu. Auch hier besteht eine ideale Möglichkeit für einen Anwendungsfall nach dem Motto: „Ask your tax department“. Dahinter steckt ein generatives Sprachmodell, über das sich Mitarbeiter aus anderen Unternehmensteilen entsprechende Infos zur steuerlichen Beurteilung eines Sachverhalts einholen können. Auch hier kann das angesammelte Wissen aus vergleichbaren Fällen in das Training des Modells mit einfließen.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in eurem Bereich voll ausschöpfen zu können?

Ein wichtiges Thema im Zusammenhang mit KI ist die Erklärbarkeit. Also die Antwort auf die Frage: Woher kommen die Informationen und wie ist das Ergebnis entstanden? Damit einher geht auch das Thema der Haftung. Denn wer haftet für ein Ergebnis, welches ausschließlich durch eine KI geliefert wurde? Wenn ein Verlag ein generatives Sprachmodell mit dem entsprechenden Verlagswissen anbietet, wird er für die Antworten bestimmt keine Haftung übernehmen. Da muss sich eine Steuerabteilung dann selbst fragen, ob das für sie ausreichend ist oder ob beim Thema Haftung wieder die Berater ins Spiel kommen sollen, die das vorgefertigte Ergebnis noch einmal fachlich überprüfen.

Daneben spielt natürlich auch das Thema Datenschutz eine entscheidende Rolle. Hiermit verbunden ist die Notwendigkeit zur Anonymisierung der Informationen, mit denen ein generatives Sprachmodell angelernt wird. Zwar sollen im Rahmen des Knowledge Sharings möglichst alle Projekte in eine KI einfließen. Dennoch muss eine bereits aus berufsrechtlichen Gründen erforderliche Trennung der mandantenbezogenen Informationen stets gewährleistet sein.

Als weiteren Punkt sind hier auch die Skills der Mitarbeiter zu nennen. Wir mögen eine Menge Ideen haben. Dennoch mangelt es teilweise schon allein an der Kapazität für die Umsetzung eben jener Ideen. Bedenkt man den demografischen Wandel, wird es in Zukunft noch schwieriger, entsprechende Mitarbeiter zu finden. Hier konnte der Steuerbereich im Vergleich zu anderen Branchen auch in der Vergangenheit nicht immer das Interesse von New Hires wecken. Daran muss unbedingt gearbeitet werden!

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Besonders große Steuerabteilungen werden sich mehr mit dem Skillset ihrer Mitarbeiter beschäftigen müssen. Dazu wird es enorm wichtig sein für die Steuerabteilung Process-Know-how zu bündeln. Ich glaube, dass es hier in Zukunft zwei Richtungen geben wird: die klassischen Tax Advisor und die Tax Prompter. Tax Prompter werden die Fähigkeit haben, der KI die richtigen Fragen zu stellen, um die richtigen Antworten zu bekommen. Dazu müssen sie in enger Zusammenarbeit mit den klassischen Steuerberatern stehen.

Aber auch die Ausbildung muss sich mit Blick auf die Zukunft verändern. Hier werden nicht mehr nur Steuerberater gefragt sein. Genauso sind Tax Prompter dafür geeignet, Schulungen zu geben und ihr Wissen weiterzugeben. Natürlich ist das auch immer abhängig von der Größe der jeweiligen Steuerabteilung. Wenn man in einer Steuerabteilung nicht für jedes Thema oder – wie in unserem Fall – für jedes Projektland einen Spezialisten in den eigenen Reihen weiß, können hier vor allem externe Berater unterstützen.

3.11 Anwendungen im Bereich Change Management

Allgemeine Informationen

In der heutigen, sich ständig verändernden Welt sind Unternehmen immer wieder mit neuen Herausforderungen konfrontiert. Diese können sich aus verschiedenen Faktoren ergeben wie bspw. neuen Technologien, veränderten Kundenanforderungen oder sich wandelnden Wettbewerbsbedingungen. Um diesen Herausforderungen zu begegnen, müssen Unternehmen entsprechende Veränderungen einleiten.

Change Management beschreibt den gesamten Prozess der Planung, Steuerung und Umsetzung von Veränderungen im Unternehmen. Ziel dessen ist es, die Veränderungsbereitschaft und -fähigkeit der Mitarbeiter zu stärken. Indem versucht wird, Veränderungsprozesse so erfolgreich wie möglich zu gestalten, wird das Change Management zu einem wichtigen Erfolgsfaktor für Unternehmen. Weil es zu einer Steigerung der Effizienz und Effektivität sowie zu einer Verbesserung der Kundenzufriedenheit und einer Stärkung der Wettbewerbsfähigkeit führen kann.

Der Erfolg von Change Management hängt dabei von verschiedensten Faktoren ab:

- › Führungskraftengagement
- › Mitarbeiterbeteiligung
- › Kommunikation und Information
- › Führungskräfteentwicklung (Führungskräfte sollten in Change Management geschult werden)

Gerade bei der Einführung und Nutzung von KI werden durch die Anpassungen im Bereich Technologie, Kundenanforderungen und Wettbewerbsbedingungen verschiedenste Veränderungen im direkten Umfeld der Mitarbeiter erforderlich. Darüber hinaus braucht es für die Nutzung von KI andere Skills und entsprechendes Know-how der Mitarbeiter. Die Einbindung von einem professionellen Change Management unterstützt dabei, den KI-Einsatz möglichst erfolgreich umzusetzen.

Experteninterview

mit Lars Bax

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret siehst du, wenn du dir die Steuerfunktion der Zukunft im Unternehmen vorstellst? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Die Einführung von KI-Technologie wird sich nicht nur auf die Steuerfunktion auswirken. Auch sämtliche administrative Funktionen im Unternehmen werden dadurch beeinflusst. Wobei die Hauptaufgabe in der Qualitätssicherung der Systeme und Prozesse liegen wird. Dann werden Systeme die entsprechenden Daten zusammenstellen, die zur Erfüllung der Compliance Aufgaben braucht.

Außerdem kann KI individuell bei der Erstellung steuerlicher Gutachten zum Einsatz kommen, um bspw. regionale Rechtsprechung zu berücksichtigen. Und auch ein Jahresabschluss auf Knopfdruck ist mittlerweile denkbar, sofern die entsprechenden Prozesse und die notwendigen Daten ordentlich strukturiert sind.

FRAGE 2: BLICK ZURÜCK

Werfen wir zunächst einen Blick zurück: Inwiefern hat das Change Management bereits in den letzten Jahren eine wichtige Rolle bei der Einführung von KI-Technologien gespielt?

Ich glaube hier muss man das Thema Change Management noch etwas breiter fassen: Auf der Reise hin zu einer neuen Technologie wird der Mensch in den Mittelpunkt gestellt, wobei die KI sicherlich einen Extremfall darstellt. Denn die Erfahrung hat gezeigt, dass selbst wenn ein logischer, effizienter und effektiver Prozess vorliegt, dass dieser nicht zwingend von den Mitarbeitern auch mit Leben gefüllt werden muss.

Erfolg stellt sich am Ende aber nur dann ein, wenn die Menschen bei der Veränderung von Anfang mit einbezogen werden und sie die genutzte Technologie durch sinnvolle Integration in die Prozesse tatsächlich akzeptieren. Grundsätzlich sollte gelten: Zuerst kommt der Mensch, dann der Prozess und dann das System.

In den letzten Jahren ist die Sensibilität der Unternehmen, die Menschen mitnehmen zu wollen, extrem gewachsen. KI muss als disruptive Änderung, ähnlich wie bei der Einführung von Internet, Mail oder Telefonie, in das Mindset der Leute eingehen. Sie müssen bereit sein, die neuen Aufgaben anzunehmen. Hier sind Unternehmen mittlerweile eher bereit Geld auszugeben.

Die Mehrheit weiß mittlerweile, dass Projekte oft nicht an inhaltlichen Themen scheitern, sondern vielmehr an der Beziehungsebene. Die wahren Gründe für ein Scheitern sind nicht selten mangelnde Ressourcen, zu gering ausfallender Managementsupport und fehlende Kommunikation. Das schließt eine mangelnde Mitnahme der Mitarbeiter natürlich mit ein. Gerade deshalb kann das Change Management an dieser Stelle eine wesentliche und vor allem unterstützende Rolle einnehmen.

Im Hinblick auf KI geht es letzten Endes um Angst. Weil sie für den Menschen schwer greifbar ist und dadurch zu Ablehnung führt. Für die Zukunft braucht es technologieoffene Menschen. Schließlich geht die Entwicklung ständig weiter. Dabei wird die technologische Hardware wie Prozessor und Taktzahl keine große Rolle mehr spielen. Vielmehr steht die Usability der Anwendungen im Fokus.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie groß ist das Bewusstsein der Mandanten für die Notwendigkeit von Change Management bezogen auf die Einführung und Nutzung von KI?

Allgemein ist die Awareness für die Notwendigkeit von Change Management bei den Mandanten in den letzten Jahren gestiegen. Natürlich in Abhängigkeit von der Branche, dem Mandanten und der Erfahrung. Die Herausforderung dabei ist, dass hier psychologische Aspekte berücksichtigt werden müssen, die nur schwer messbar sind. Da die Steuerbranche sehr zahlenaffin ist, ist der Mehrwert teilweise noch schwer zu vermitteln.

Doch wir sehen immer wieder, dass Projekte an den gleichen Dingen scheitern. Zu diesen gehören Überforderung bei der Umsetzung von Projekten zusätzlich zur alltäglichen Arbeit oder Führungskräfte, die nicht abgeholt wurden. Eine weitere Problematik stellt auch eine fehlende Management Attention dar, wenn diese schon ans nächste Projekt denken. Hier brauchen wir Change Management, um auf Basis einer Analyse der Ängste und Probleme der Menschen im Unternehmen Konzepte zu entwickeln. In erster Linie, um die psychologische Ebene bespielen zu können. Es ist damit die Professionalisierung von vielen kleinen Dingen, die man im Projektmanagement eher nebenbei macht.

Aufgrund der hohen Fluktuation am Arbeitsmarkt und der damit verbundenen Schwierigkeit von Nachbesetzungen, ist der Fokus auf die eigenen Mitarbeiter größer geworden. In der Folge ist auch die Awareness für das Thema Change Management gestiegen. Mittlerweile kennt jede Führungskraft das Konzept, wendet es allerdings operativ eher selten an.

FRAGE 4: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI im Steuerbereich voll ausschöpfen zu können?

Die größte Herausforderung für die effektive Nutzung von KI ist immer noch der Mensch selbst. Denn ein KI-Modell ist immer nur so gut wie der Nutzer, der die Maschine bedient. Die Job-Beschreibung vieler Mitarbeiter, die früher Experten auf ihrem Sachgebiet waren, wird sich durch den Einsatz von KI-Technologien ändern. Beispielsweise wird sich ein Experte für Excel-Anwendungen zukünftig mehr und mehr mit dem Formulieren von Prompts beschäftigen müssen, da Excel-Kalkulationen von KI-Systemen effektiver ausgeführt werden können.

Gerade beim sogenannten Prompting wird es eine neue Art von Kreativität brauchen. Neben den Experten für die Steuerfunktion werden sich viele Mitarbeiter zu KI-Experten weiterentwickeln müssen, die auch die Logik hinter den Modellen verstehen.

Auch die eingangs erwähnte Qualitätssicherung wird eine entscheidende Rolle spielen, um die Ausgaben der KI-Systeme als richtig oder falsch zu bewerten. Betrachtet man diese anstehenden Veränderungen in den jeweiligen Berufsbildern, muss auch die Ausbildung dahingehend angepasst werden.

Eine weitere Herausforderung bei der Nutzung von KI ist das Thema Datenschutz. Dieses wird häufig als Argument missbraucht, um Dinge nicht tun zu müssen. Hier brauchen wir mehr Lösungsorientiertheit. Trotzdem ist es wichtig, sich mit Datenschutz zu beschäftigen. In erster Linie um auszuloten, was möglich ist und was nicht.

FRAGE 5: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Je technologischer die Steuerfunktion wird, desto mehr Leute mit einem Verständnis für die Technologie und End-to-End Prozesse werden benötigt. Zudem müssen sie in der Lage sein, die genutzten Daten auch tatsächlich verstehen. Weil viele von diesen Daten gar nicht in der Steuerabteilung entstehen.

Als Berater können wir zum einen dabei unterstützen, die richtigen Personen für diese Aufgaben zu finden und auch eine entsprechende Ausbildung voranzutreiben. Zum anderen können wir aber auch weiterhin die Steuerexperten mit Fachwissen zur Verfügung stellen.

Einen der Grundpfeiler stellt das sogenannte Lean-Management dar – also die menschenorientierte Prozessoptimierung. Wir wollen ein selbstlernendes Team sein, das sich in kleinen Schritten optimiert und Veränderungen durch Erfolge positiv wahrnimmt. Durch die Schaffung einer entsprechenden Atmosphäre kann so ein Raum für Lean-Thinking entstehen. Meine generelle Empfehlung: Führungskräfte sollten als Coach auftreten und eine Atmosphäre schaffen, in der die Leute kreativ sein können, Zeit und Ressourcen zur Verfügung gestellt bekommen, um Dinge ausprobieren zu können und dies auch wertgeschätzt wird.

FRAGE 6: BLICK IN DIE ZUKUNFT

Welchen Nutzen von KI siehst du im Bereich Change Management selbst? Wie kann KI hier aktuell und in Zukunft bei der täglichen Arbeit sinnvoll unterstützen?

In unserem Bereich gibt es ein Team, das sich besonders mit Daten und Technologie beschäftigt. Hier können KI-Lösungen zur Datenstrukturierung verwendet werden. Alternativ können sie aber auch zur Einordnung und Auswertung von Interviews zum Einsatz kommen, die wir im Rahmen unserer Arbeit führen. Gleiches gilt für das Thema Wissensmanagement mit in-house Know-how, bei dem KI-Anwendungen wie bspw. ChatGPT eine willkommene Unterstützung darstellen.

Grundsätzlich lassen sich Anwendungsmöglichkeiten überall dort finden, wo Daten im Spiel sind. Generell gilt, dass wir einen großen Fokus auf Individualität und den Menschen legen. Daher ist für uns ein Gespür für dessen Belange und Anforderungen wichtiger ist als die Nutzung von KI. Getreu dem Motto: „Es geht nicht ohne den Menschen und wenn wir es nicht richtig machen, auch nicht mit“.

3.12 Anwendungen im Bereich FS Tax

Allgemeine Informationen

Der Begriff Finanzindustrie umfasst als übergreifendes Synonym für die Geld- und Finanzbranche alle finanziellen Dienstleister. Also nicht nur Banken, Versicherungen, Fonds, usw. Ebenfalls gemeint sind geschäftsvermittelnde und -begleitende Intermediäre, Kontrollinstitutionen, Ratingagenturen und viele mehr.

Dreh- und Angelpunkt dieser monetären bzw. finanziellen Institutionen und Prozesse sind letztlich Daten. Die Finanzindustrie ist in ein dichtes regulatorisches Netzwerk eingebettet. Hierzu zählen auch die nationalen und internationalen Steuerbehörden. Diese bedienen sich der Finanzindustrie als Datenlieferant, fungieren als Kontrollinstanz und übernehmen der Finanzindustrie substantielle Steuerprozesse wie z. B. im Bereich der Kapitalertragsteuern.

Experteninterview mit Robert Welzel

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret siehst du, wenn du dir die Steuerfunktion der Zukunft im Unternehmen vorstellst? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Die Steuerabteilung befindet sich aktuell in einem notwendigen Transformationsprozess. Dabei stellt sich aktuell die Frage: Sind die Steuerfachexperten überhaupt die richtigen Leute, um den Wandel voranzutreiben? Oder begleiten sie diesen eher nur am Rande? Denn hier sind vor allem Prozessfachleute gefragt, die wissen wie Abläufe gesteuert werden können. Diese Notwendigkeit zeigt sich auch an den Universitäten, wo eine neue Generation in einem Dreiklang aus Accounting, Programmierkenntnissen und Steuerfachwissen ausgebildet wird. In erster Linie um die Anforderung der Steuerabteilung der Zukunft abbilden zu können.

Dabei geht es im Bereich Financial Services nicht nur um die Frage der Optimierung von Prozessen mit ChatGPT und Prompts. Es geht auch um die Frage, welche neuen Geschäftsmodelle durch die KI entstehen werden und welche steuerlichen Anforderungen sich daraus entwickeln. Schon heute existieren viele Regulierungen wie z. B. DAC6, AML, ATAD oder CARF, durch die Daten abgefragt werden und die zu einer hohen Datengranularität führen. Hier liegt eine zentrale Aufgabe der Steuerfunktion der Zukunft. Sie muss die steuerlichen Daten mit Hilfe von KI sinnvoll in Datenmodellen abbilden und analysieren, um alle möglichen Stakeholder bedienen zu können. Zusammengefasst heißt das: Komplexität beherrschbar machen und mögliche Value-adds erschließen.

Für die Zukunft erwarte ich, dass die Bedeutung von KI im Bereich Financial Services noch viel stärker zunehmen wird. Dabei werden sich neue Geschäftsmodelle entwickeln, die aus Compliance-Perspektive begleitet werden müssen und neue aufsichtsrechtliche Fragestellungen hervorrufen. Durch Tokenisierung von Digital Assets können weitere Komponenten wirtschaftlicher Wertschöpfungsprozesse monetarisiert und dem Finanzmarkt als neue Assetklasse zur Verfügung gestellt werden. Existierende Strukturen und Vertriebsmodelle sowie Anforderungen an Compliance werden sich dadurch komplett verändern.

FRAGE 2: BLICK ZURÜCK

Wie schätzt du den Mehrwert von KI mit Blick auf die vergangenen Jahre in deinem Bereich ein – also die Zeit vor der Ära der generativen KI?

Auch in der Vergangenheit gab es schon verschiedene KI-Projekte, meistens im Bereich des Machine Learning. Dabei ging es z. B. um die Optimierung von Aktiengewinnen oder die Nutzbarmachung von Verlusten, indem Finanzinstrumente eingesetzt oder Trade Patterns im Kontext von Quellensteuern analysiert werden. Der Umfang und die Komplexität der Datengrundlage ist dabei so groß, dass eine Analyse mittels Excel und Pivot-Tabellen fast nicht mehr möglich ist. Hier konnte man mit Hilfe der KI positive Resultate erzielen.

Ein anderes Beispiel für den Einsatz von KI war im Rahmen der Klassifizierung von Finanzinstrumenten für das Steuerrecht. Eine manuelle Analyse verursacht hierbei hohe Kosten, sodass man durch den Einsatz von KI sehr gute Erfahrungen machen konnte. Und auch für die Zukunft ist zu erwarten, dass die Klassifikation mittels KI sich weiter verbessern wird.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt du aktuell das Potenzial für KI in deinem Bereich ein und welchen Impact haben deiner Ansicht nach generative Sprachmodelle?

Die Nachfrage und das Interesse an KI ist aktuell sehr groß. Und alle haben gewisse Vorstellungen, was in der Zukunft noch passieren wird. Doch für konkrete Lösungen sehe ich bei vielen Mandanten derzeit noch eine gewisse Zurückhaltung. Hier wollen wir als Berater unsere Mandanten unterstützen, um Prioritäten herauszufinden und sie auf dem Weg der Entwicklung zu begleiten.

Und auch die Finanzverwaltung digitalisiert sich mehr und mehr. Wobei das Ausland tendenziell schon einen Schritt weiter ist als unsere heimische Finanzverwaltung. Hier ist es wichtig ein Netzwerk zu bilden mit anderen Serviceprovidern und Softwareentwicklern gemeinsam an Lösungen zu arbeiten und diese auch mit der Finanzverwaltung zu teilen. Dabei braucht es nicht nur abstrakte Anwendungsmodelle, sondern vor allem greifbare Prototypen.

FRAGE 4: ANWENDUNGSFÄLLE

Bei welchen konkreten Prozessen in deinem Bereich siehst du aktuell das größte Potenzial zur Nutzung von KI, besonders im Hinblick auf aktuelle technologische Neuerungen wie bspw. generative Sprachmodelle?

Wenn im Bereich Financial Services Investoren aus unterschiedlichen Ländern über intermediäre Strukturen (z. B. Fonds) in ein breit gestreutes Asset-Universum aus unterschiedlichen Ländern investieren, dann bewirkt dies eine hohe steuerliche Komplexität. Was wäre zum Beispiel steuerlich zu beachten, wenn ein Investor aus Frankreich in ein Asset in Indien investiert? Hier kann KI eingesetzt werden, um solche komplexen Strukturen mit unterschiedlichen involvierten Ländern beherrschbar zu machen. Dabei geht es vor allem um die Verknüpfung von strukturierten Investitionsdaten mit regulatorischen und steuerlichen Anforderungen. Eine effiziente Durchführung solcher Analyseprozesse kann dabei einen spürbaren Mehrwert schaffen.

In Zukunft wird man ein Asset lediglich ins System eingeben und die KI gibt an, wie es für welchen Investor in welchen Ländern besteuert wird. Dabei ist davon auszugehen, dass Finanzinstitute ihren Anlegern diese Informationen in Zukunft automatisch zur Verfügung stellen müssen. In diesem Bereich haben wir mit der Tax Risk Radar Function bereits eine erste Lösung für Fondsgesellschaften aufgebaut. Hierbei wird geprüft, in welchen Ländern die Fonds investiert sind, wie dessen Assets zu klassifizieren sind und was sich daraus für Risiken in den einzelnen Ländern ergeben.

Vor dem Hintergrund der rasanten aktuellen Entwicklung der generativen Sprachmodelle und abhängig von deren Training mit Landesspezifika, kann dabei nicht mehr nur eine statische Analyse erfolgen, sondern zukünftig auch eine dynamische Analyse. Dabei führt die Verbindung von Transaktionsdaten mit juristischen und steuerrechtlichen Fragestellungen zu einem verbesserten Tax Risk Monitoring.

FRAGE 5: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in deinem Bereich voll ausschöpfen zu können?

Die grundlegende Frage ist, ob man selber technologieaffin genug ist, um die neue Entwicklung sinnvoll begleiten zu können. Denn viele der KI-Modelle sind sehr mathematisch und formallastig, sodass man schnell an Grenzen stoßen kann. Heißt konkret: Man muss etwas vorantreiben, das man selbst vielleicht nicht zu 100% versteht. Dadurch können dann Akzeptanzprobleme auftreten, die eine erfolgreiche Einführung erschweren.

FRAGE 6: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

In der Zukunft wird unwiderruflich ein „War of Talent“ stattfinden. Schon jetzt lernen Studenten an den Universitäten einen Mix aus IFRS-Kenntnissen, Programmierfähigkeiten und Steuerkenntnissen. Genau nach diesen Personen werden Unternehmen suchen. Doch die eigentliche Frage dabei wird sein: Welche Neugierde kann man bei Personen hervorrufen, damit sie sich für den Steuerbereich interessieren?

Aber auch bei den aktuell angestellten Mitarbeitern gibt es einige Personen mit großem Interesse an Weiterentwicklung. Diese Leute müssen gefördert werden, bspw. durch Kurse im Internet. Hier bestehen bereits zahlreiche Möglichkeiten, Zertifikate bei Microsoft, Google oder AWS zu erwerben.

Wie der Berater erfolgreich unterstützen kann, zeigt das Beispiel der Betriebsprüfung. Hier zeigen Analysen mittlerweile schnell, wenn eine Steuerarbitrage stattgefunden hat. Dazu wird sich die Datentransparenz nach §42AO oder §39AO weiter zu Lasten des Steuerpflichtigen verschieben. Die eigentliche Kernaufgabe besteht also darin, die vielen Daten zu visualisieren und mit der Reproduktion von möglichen Analysen im Rahmen des Tax Risk Managements einen Mehrwert zu erzeugen.

3.13 Anwendungen im Bereich Tax Allgemein - Of Counsel Hans Maier

Allgemeine Informationen

Seit 2021 unterstützt Dr. Hans Maier die WTS als Of Counsel am Standort Stuttgart. Der Rechtsanwalt und ehemalige Global Head of Tax der Robert Bosch GmbH verfügt über mehr als 30 Jahre Erfahrung im Management von Steuerfunktionen multinationaler Unternehmen. Als einer der bekanntesten Steuerchefs in Deutschland hat er bei Bosch eine leistungsstarke globale Steuerfunktion mit einem weltweit agierenden Team und umfassenden digitalisierten steuerlichen Prozessen aufgebaut.

Seine Karriere startete Dr. Hans Maier im Jahr 1989 als Referent für internationales Steuerrecht bei Bosch. Nach verschiedenen weiteren Positionen im In- und Ausland zeichnete er sich dort ab 2005 weltweit für die Bereiche Steuern, Zölle und Exportkontrolle verantwortlich. Bestens vertraut mit steuerlichen Aufbau- und Ablauforganisationen, weiß er als Konzernpraktiker, wie Steuer- und Zollprozesse in großen Organisationen effizient, Compliance-sicher und digitalisiert aufgesetzt werden.

Experteninterview mit Hans Maier

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

Was konkret siehst du, wenn du dir die Steuerfunktion der Zukunft im Unternehmen vorstellst? Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?

Die Steuerfunktion befindet sich aktuell in einer Transformation, wobei das Thema Digitalisierung nur eines der immens wichtigen Felder ist. Hier glaube ich, dass sich das Anforderungsprofil an die Mitarbeiter in Zukunft dramatisch ändern wird. Ich bin überzeugt davon, dass vermehrt „mixed Teams“ zum Einsatz kommen werden, bestehend aus Steuerberatern und Leuten mit IT und Prozessfokus. Doch da sich das Berufsbild insgesamt ändert, müssen Steuerberater sich zwangsläufig auch digitale Kompetenzen aneignen. Durch den Einsatz von KI wird sich die Rolle des klassischen Steuerberaters hin zu einem steuerlichen Qualitätsmanager weiterentwickeln.

Viele Unternehmen stellen sich organisatorisch entsprechend der End-to-end Prozesse auf. Hier muss sich die Steuerabteilung fragen, wie sie einen Wertbeitrag leisten kann. Dies erfordert eine stärkere Zuwendung zu den ERP-Systemen, die in Abstimmung mit der gesamten IT-Strategie eines Unternehmens erfolgen muss. Hier befinden sich viele Steuerabteilung aufgrund fehlender Kompetenzen noch im „Dornröschenschlaf“, wodurch sie im Unternehmen zum Bremser wird und ihr Stellenwert entsprechend sinkt.

Was den Bewerbermarkt angeht, sucht man häufig vergebens nach einer eierlegenden Wollmilchsau mit IT-, Prozess- und Steuerfachkenntnissen. Doch ein guter IT'ler kann sich auch anderes Wissen aneignen. Dafür braucht es dann allerdings das nötige Vertrauen, jemanden einzustellen, der von Steuern keine Ahnung hat und intern weiterentwickeln werden kann. Hier liegt auch ein großes Potenzial für die Beratung, denn der Mittelstand kann dem nicht gerecht werden.

Und nicht zuletzt wird der seit Jahren zu beobachtende Trend der Digitalisierung auch die Finanzverwaltung erreichen. Das lässt sich unweigerlich am Thema Real Time Reporting erkennen. Einige Finanzverwaltungen wie Mexiko, Brasilien oder Osteuropa sind hier sogar schon einen Schritt weiter und mittlerweile bereits in der Lage, viele Daten auszuwerten.

FRAGE 2: BLICK ZURÜCK

Wie schätzt du den Mehrwert von KI mit Blick auf die vergangenen Jahre in eurem Bereich ein – also die Zeit vor der Ära der generativen KI?

In den vergangenen Jahren gab es eine steile Lernkurve was das Thema KI angeht. Dabei hat der Einsatz häufig nur dann Sinn gemacht, wenn man eine große Anzahl von Daten hatte, die zunächst strukturiert werden mussten. Bei Bosch wurde im Jahr 2018 eine eigene KI-Truppe aufgebaut, die sich speziell auf die Suche nach Anwendungsfällen gemacht hat und direkt an den Head of Tax berichtete. So wurde KI unter anderem bei der SAP VAT Compliance eingesetzt, der Plausibilisierung im Rahmen der Umsatzsteuer oder aber auch im Zollbereich der der Tarifierung.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt du aktuell das Potenzial für KI in deinem Bereich ein und welchen Impact haben deiner Ansicht nach generative Sprachmodelle?

Aus meiner Sicht wird KI die Steuerfunktion weiterhin dramatisch verändern. Zwar hat ChatGPT seit seiner Einführung einen großen Hype ausgelöst und das Thema KI näher zu den Leuten gebracht. Doch aufgrund mangelnder Vertraulichkeit ist es keine valide Option für viele Anwendungsfälle im Steuerbereich.

Eine Eigenentwicklung von generativen Sprachmodellen ist hingegen relativ teuer. So lässt sich festhalten, dass die Digitalisierung der Steuerfunktion länger dauert als gedacht, was durchaus zu Frust führen kann. Dabei ist es enorm wichtig, ein vernünftiges Expectations Management zu betreiben, was mit der KI möglich ist und was rechtssicher umgesetzt werden kann. Aus meiner Sicht gibt es nichts Schlimmeres als einen Berater, der etwas verspricht und im Nachhinein dann nichts davon halten kann.

FRAGE 4: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in deinem Bereich voll ausschöpfen zu können?

Als Herausforderung im Zusammenhang mit der KI sehe ich vor allem die Erklärbarkeit und Dokumentation der Ergebnisse aus einer KI. Daneben spielen aber auch betriebswirtschaftliche Fragestellungen eine Rolle. Soll heißen: Wie viel Geld muss ich einsetzen und was für einen Nutzen bringt es mir. Grundsätzlich muss der Kosteneinsatz überschaubar bleiben und die Vorteile müssen höher ausfallen. Das sind zwei wesentliche Punkte, an denen bereits viele Projekte in der Vergangenheit gescheitert sind.

In diesem Zusammenhang sind natürlich auch die Themen Datenschutz und Vertraulichkeit absolut zentral: Wo gehen meine unternehmensinternen Daten hin, wenn ich eine KI damit füttere? Sensible Daten dürfen die Sphäre des Unternehmens nicht ungeschützt (oder „nur maximal geschützt“) verlassen, sonst kommt eine solche Technologie niemals zum Einsatz.

FRAGE 5: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Als Steuerabteilung braucht man Mitarbeiter mit den entsprechenden Kompetenzen und eine Organisation, welche die Transformation unterstützt und auch einfordert. Darüber sollte immer eine klare Vision und Strategie stehen. Betrachtet man den Compliance-Bereich, so entwickelt dieser sich mehr und mehr hin zu einem non-core-Bereich, dessen Bedeutung zurückgeht. Diesen kann man zwar inhouse abbilden, muss es mit der richtigen Beraterunterstützung von außen aber nicht.

Die zentrale Steuerfunktion hat als Kernaufgabe auch immer eine weltweite Ordnungs- und Risk-Management-Funktion. Dafür ist es wichtig, auch die Kompetenzen in den anderen Ländern einer Gruppe zu nutzen und sich intensiv mit den ERP-Systemen sowie internen Prozessen auseinanderzusetzen.

Die Rolle des Beraters kann dabei sehr vielfältig ausfallen. Größere Unternehmen wie Bosch sind in der Entwicklung schon deutlich weiter als der Mittelstand und benötigen demnach andere Unterstützung seitens der Berater. Bei größeren Firmen sollte der Berater vor allem beim Projektmanagement unterstützen und Hinweise bzw. Anregungen für die Strategie geben. Beim Mittelstand hingegen sollte der Berater als Coach für den Head of Tax agieren. Dabei ist die Frage zu klären, wie man die Steuerfunktion digitalisieren kann. Lediglich ein Tool zu kaufen, ist hier nämlich nicht ausreichend und führt schnell zu Frustration.

3.14 Anwendungen im Bereich Tax Allgemein - Of Counsel Brigitte Fischer

Allgemeine Informationen

Seit 2023 unterstützt Brigitte Fischer die WTS als Of Counsel am Standort in München. Die langjährige Steuerchefin der Freudenberg Gruppe ist für ihre umfangreiche Erfahrung bekannt, insbesondere bei der nationalen und internationalen Besteuerung von Personengesellschaften.

Ihre Karriere begann Brigitte Fischer in der Beratungsgesellschaft Deloitte in Düsseldorf. Im Jahre 1992 wechselte sie als Steuerreferentin zur Freudenberg Group nach Weinheim, wo sie ab 1998 die Position des Head of Corporate Tax and Customs einnahm und für 25 Jahre bekleidete. Dadurch ist sie bestens vertraut mit den steuerlichen Prozessen multinationaler Konzernen und den damit verbundenen Herausforderungen.

Experteninterview mit Brigitte Fischer

FRAGE 1: STEUERFUNKTION DER ZUKUNFT

*Was konkret siehst du, wenn du dir die Steuerfunktion der Zukunft im Unternehmen vorstellst?
Wie kann diese künftig einen Mehrwert im Unternehmen generieren und wie werden die technologischen Innovationen deren Arbeit im Unternehmen in den nächsten Jahren verändern?*

Auch in der Zukunft werden gewisse Aufgaben der Steuerabteilung gleichbleiben: hierzu zählen u. a. Deklarationsaufgaben, Betriebsprüfungen, steuerliche Begleitung von M&A-Prozessen oder die Erfüllung von Compliance-Vorschriften. Doch bei der Wahrnehmung dieser Aufgaben wird es mehr technische Unterstützung geben. Und damit viele Aufgabenbereiche ersetzen, die bislang zumeist vom Mitarbeiter in der Steuerabteilung bearbeitet wurden. Dafür ist es natürlich notwendig, dass die Mitarbeiter mehr digitale Expertise aufbauen. Was in der Vergangenheit nicht immer Priorität in den Steuerabteilungen hatte. Durch die neuen Technologien wird steuerliches Detailwissen weniger wichtig, da es leicht und schnell mittels KI generiert werden kann. Dafür steigt die Bedeutung, die steuerliche Systematik (national wie international) zu verstehen, um von der KU bereitgestellten Ergebnisse verifizieren zu können.

Im Unternehmen werden die Steuerabteilungen oft als Bremser und Bedenkenträger wahrgenommen. Dies hat zur Folge, dass die Steuerabteilungen häufig relativ spät oder manchmal gar nicht in relevante Businessentscheidungen einbezogen werden. Hier muss sich auch die Haltung der Steuerabteilungen dringend ändern. Indem man mehr auf die Bedürfnisse des operativen Geschäfts des Unternehmens eingeht. Analog zur Maßgabe: „Steuern sind wichtig, aber sie sind nicht das Wichtigste im Unternehmen“.

Inwiefern KI hierauf einen Einfluss nehmen wird, bleibt abzuwarten. Allerdings besteht durch generative Sprachmodelle die Möglichkeit, für die operativ Verantwortlichen in den Geschäftsgruppen steuerliche Informationen zu beschaffen, bevor die Steuerabteilung direkt kontaktiert wird.

Und auch im Bereich der Lehre wird sich in Zukunft einiges ändern müssen. Aufgrund des Fachkräftemangels im Recruiting ist es heutzutage ohnehin schon überaus schwer, IT-affine Mitarbeiter mit Steuerexpertise zu finden. Allein dadurch ist die intensive Zusammenarbeit von Unternehmen und Hochschulen umso wichtiger, wenn diese künftig einen gewissen Einfluss auf die Lehre nehmen wollen.

FRAGE 2: BLICK ZURÜCK

Wie schätzt du den Mehrwert von KI mit Blick auf die vergangenen Jahre in deinem Bereich ein – also die Zeit vor der Ära der generativen KI?

In den vergangenen Jahren hat die KI im Steuerbereich noch kaum eine Rolle gespielt. Lediglich in Rahmen von Übersetzungen wurden hier bereits Tools eingesetzt, die dabei schon einen Vorgeschmack auf die Möglichkeiten und die Qualität von KI gegeben haben. Die Zurückhaltung bei dem Thema mag auch damit zusammenhängen, dass viele gedacht haben, dass die KI-Entwicklung nicht in der jüngst zu beobachtenden Geschwindigkeit kommen wird. Die Affinität für KI muss in ein Unternehmen reingebracht werden, damit diese sich in diesen Punkten überhaupt weiterentwickeln kann. Dabei sollte immer auch auf andere Abteilungen wie bspw. den Business-Services-Bereich geschaut werden, in denen der KI-Einsatz mit der Nutzung von Chatbots bereits weiter fortgeschritten ist.

FRAGE 3: BLICK AUF DIE GEGENWART

Wie schätzt du aktuell das Potenzial für KI in deinem Bereich ein und welchen Impact haben deiner Ansicht nach generative Sprachmodelle?

Ich schätze das Potential in vielen Bereichen als hoch ein, z. B. im Bereich von Recherchetätigkeiten oder beim Abfassen einfacherer steuerlicher Stellungnahmen. Das gilt aber auch für die Bearbeitung von steuerlichen Massenprozessen (USt, Zoll, Quellensteuern) und im Bereich von TP. Ohne Frage, die Nutzung von generativen Sprachmodellen wie ChatGPT erfordert auch in Steuerabteilungen ein ziemliches Umdenken. Ich denke daher, dass die KI-Nutzung von höchster Ebene in den Unternehmen eingefordert und vorangetrieben werden muss.

Hierfür müssen allerdings die Bereitschaft und der Wille zur Nutzung von KI vorhanden sein und die entsprechenden finanziellen Mittel bereitgestellt werden. Denn von den Mitarbeitern alleine wird die Entwicklung nicht vorangetrieben werden können. Für die Steuerfunktion in einem Unternehmen ist es dabei wichtig, Mitarbeiter in der Steuerabteilung zu beauftragen, sich intensiv mit den Potenzialen von KI auseinanderzusetzen. Denn diese sind zweifelsfrei vorhanden.

FRAGE 4: HERAUSFORDERUNGEN

Worin bestehen aktuell die größten Herausforderungen, um das Potenzial von KI in deinem Bereich voll ausschöpfen zu können?

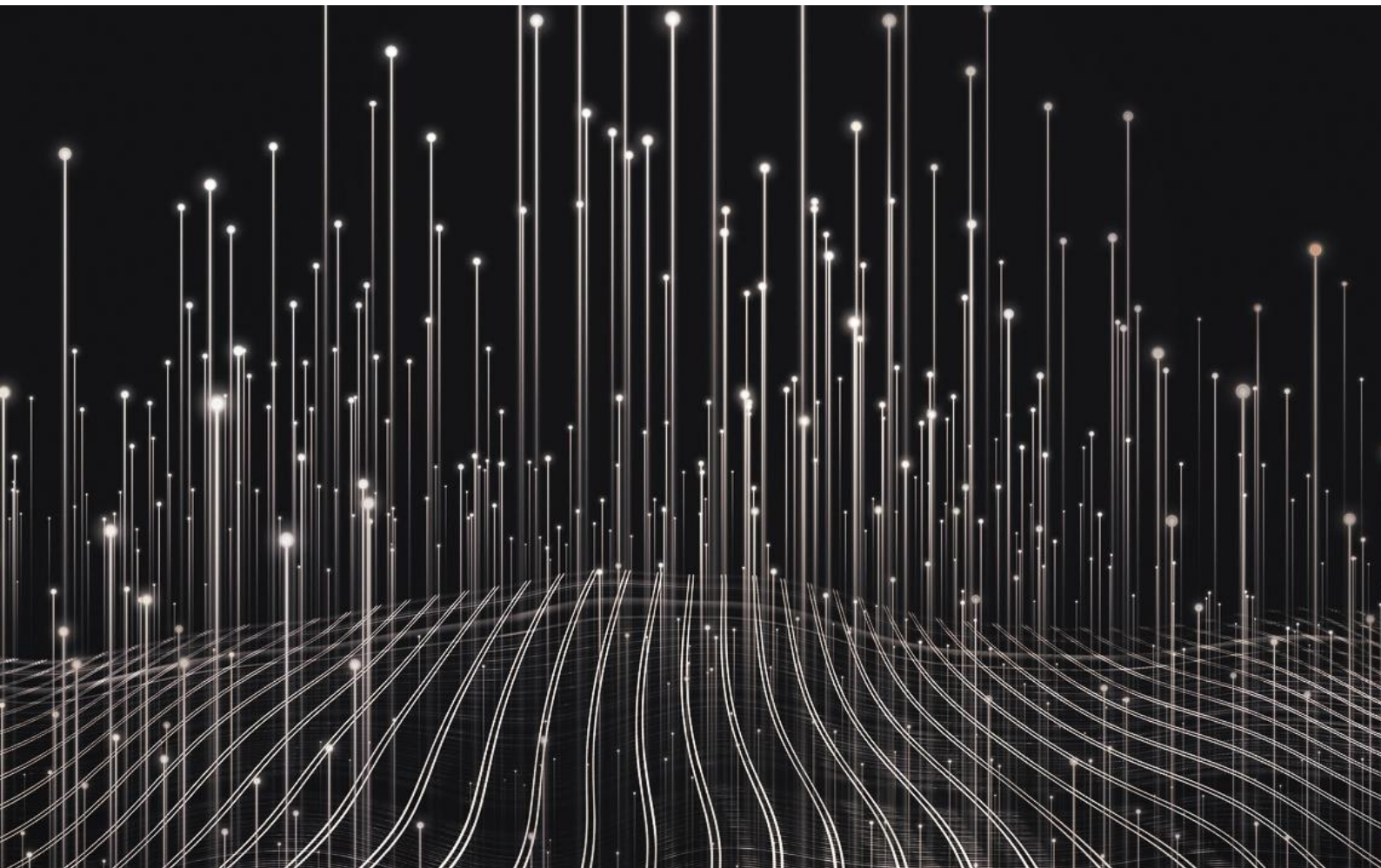
Ich glaube, es ist die Unkenntnis. In vielen Bereichen wird nicht gesehen, was durch den Einsatz von KI möglich ist und welche Arbeitserleichterungen damit erreicht werden können. Aber natürlich spielt auch das liebe Geld eine Rolle. Denn nicht jeder ist auch dazu bereit, in das Thema zu investieren. Und zuletzt braucht man auch die entsprechenden Leute in den Steuerabteilungen, die sich dem Thema annehmen, die die Prozesse im Unternehmen kennen und entsprechende Einsatzmöglichkeiten identifizieren können. Das müssen nicht alle Mitarbeiter einer Steuerabteilung sein. Aber zwei bis drei braucht es einfach, um das Thema voranzutreiben.

FRAGE 5: AKTUELLER HANDLUNGSBEDARF

Angesichts der schnell voranschreitenden technologischen Innovation, wie müssen sich Steuerabteilungen aktuell aufstellen und welche unterstützende Rolle können Berater hier übernehmen?

Die Steuerabteilungen müssen Leute bekommen, die sich neuen Technologien gegenüber öffnen wollen und auch können. Dabei besteht auch die Möglichkeit, in der IT-Abteilung Leute zu finden, die eine Affinität für Steuern haben. Denn generell wird das Thema des Fachkräftemangels die Zukunft prägen. Wenn intern keine Kapazitäten vorhanden sind, bleibt den Unternehmen nichts Anderes übrig, als sich weitere Unterstützung von den Beratern zu holen.

Als Fritz Esterer vor 5 Jahren mit dem Thema KI zu Freudenberg kam, war er seiner Zeit weit voraus. Doch heutzutage kommt eigentlich kein Unternehmen mehr drum herum, sich von Beratern die Möglichkeiten von KI aufzeigen zu lassen. Das gilt gerade auch für kleinere und mittelgroße Unternehmen, die im Vergleich zu den großen Unternehmen nicht die nötigen Ressourcen in den Steuerabteilungen haben. Doch auch die großen Unternehmen werden Probleme im Recruiting bekommen und dadurch gezwungen sein, fehlende Personalkapazitäten durch den Einsatz von Künstlicher Intelligenz abzufedern. Ein komplettes Abfangen des Personalmangels dürfte mit Unterstützung durch die KI aber wohl kaum gelingen.



4 Anwendungsszenarien

4.1 Umsatzsteuerliche Klassifizierung von Leistungsbeschreibungen von Rechnungen

Thema & Problemstellung

Nach § 14 Abs. 4 Satz 1 Nr. 5 UStG muss eine Rechnung zur Berechtigung des Vorsteuerabzugs eine Leistungsbeschreibung enthalten. Diese soll es ermöglichen, die Leistung eindeutig zu identifizieren und leicht nachprüfbar zu machen. Aufgrund einer zu allgemeinen Formulierung, dem Mangel an Details oder branchenspezifischer Besonderheiten kann es jedoch zur Ablehnung des Vorsteuerabzugs kommen.

Prinzipiell werden mit dem Schreiben vom 01.12.2021 des Bundesministeriums für Finanzen (BMF)¹ die inhaltlichen Voraussetzungen für die Bezeichnung einer Leistung in einer Rechnung näher definiert. Dennoch ist nach wie vor nicht eindeutig bestimmt, was eine zutreffende Leistungsbeschreibung tatsächlich ist. Der Interpretationsspielraum, ob bspw. eine Ware in der Leistungsbeschreibung als „handelsüblich“ beschrieben wurde, stellt insbesondere Steuerabteilungen in Unternehmen vor eine große Herausforderung. So gelten Lieferungen mit Gattungsbezeichnungen wie „T-Shirt“, „Bluse“ o. ä. als handelsüblich. Demgegenüber ist bei sonstigen Leistungen keine entsprechende Angabe vorgesehen. So wird z. B. die Leistungsbeschreibung „Erbringung juristischer Dienstleistungen“ von der Finanzverwaltung als eine nicht ausreichende Leistungsbeschreibung beurteilt².

Da die Kontrolle der Leistungsbeschreibung komplex und häufig noch manuell erfolgt, befasst sich vorliegendes Anwendungsszenario mit der Frage, inwieweit sich große Sprachmodelle für die Klassifikation von Leistungsbeschreibungen nutzen lassen können. Ähnlich einem Vier-Augen-Prinzip soll durch die KI eine Vorabprüfung stattfinden, sodass der Steuerexperte nur noch solche Leistungsbeschreibungen manuell prüfen muss, die tendenziell nicht für einen Vorsteuerabzug ausreichend sind. Aufgrund des repetitiven Charakters des Prüfprozesses, aber auch der Quantität an zu prüfende Rechnungen, birgt dieser Ansatz ein enormes Automatisierungspotenzial, welches den Steuerexperten in dessen täglichen Arbeit maßgeblich entlasten können.

Lösungsarchitektur

Ziel des Anwendungsszenarios ist die Bewertung, inwieweit sich große Sprachmodelle mit ihrem vorhandenen allgemeinen Wissen und ihrer Generalisierungsfähigkeit für die Klassifikation von Leistungsbeschreibungen nutzen lassen. Für die Anpassung eines LLM an eine spezifische Aufgabe gibt es unterschiedliche Ansätze, die in der Literatur diskutiert werden [40]. Die Lösungsarchitektur des Anwendungsfalls basiert auf unterschiedlichen Prompt-Tuning-Ansätzen.

Hard Prompt Tuning: Der Fokus liegt weniger auf der Bereitstellung von Kontextinformationen, sondern auf der Ermittlung der besten Prompt-Formulierung. Durch eine möglichst präzise Formulierung soll dem Sprachmodell wenig Interpretationsspielraum gegeben werden, um so eine möglichst präzise Antwort zu erhalten. Entsprechend wurden unterschiedliche Formulierungen verprobt und evaluiert.

In-Context Learning (ICL): Die Idee besteht darin, Beispiele und zusätzliche Informationen zur Lösung der spezifischen Aufgabe bereitzustellen. Dadurch ist das Sprachmodell in der Lage, das gewünschte Verhalten und Informationen abzuleiten und entsprechende Antworten zu generieren. Der Vorteil liegt darin, dass für diese Art des Fine-Tunings nur eine begrenzt oder gar keine gelabelten Trainingsdaten verfügbar sein müssen. Außerdem lässt sich der Ansatz schnell implementieren und auch perspektivisch für andere

¹ Bundesministerium der Finanzen (BMF), 1.12.2021, III C 2 - S 7280 - a/19/10002 :001

² BFH-Urteil vom 10. 11. 1994, V R 45/93, BStBl 1995 II S. 395

steuerliche Aufgaben nutzen, bei denen ein Fine-Tuning der Modellparameter nicht möglich ist, z. B. aufgrund fehlendem Direktzugriff auf das Sprachmodell.

Few-shot Prompt-Tuning: Bei diesem verprobten Verfahren handelt es sich um eine Methode, im Zuge derer das LLM auf eine spezifische Aufgabe anhand einer sehr begrenzten Anzahl von Beispielen angepasst wird – den sogenannten „Shots“.

Die Umsetzung des Anwendungsfalls erfolgte in Form eines binären Klassifikationsproblems auf den zwei LLMs *LeoLM*³- und *GPT-4*⁴.

LeoLM: Dabei handelt es sich um ein auf der deutschen Sprache optimiertes Llama-2-Modell, basierend auf einer Transformer-Architektur [41]. Das auf Chat-Texten abgestimmte Modell wurde mit dem Ziel ausgewählt, überwiegend in deutscher Sprache vorliegende Anfragen und Leistungsbeschreibungen besser zu verstehen und möglichst kohärente, relevante und kontextuell konsistente Antworten zu generieren. Die Implementierung des Anwendungsfalls erfolgte innerhalb der internen DFKI-Infrastruktur. Die Lizenzierung dieses Modells erlaubt eine kommerzielle Nutzung.

GPT-4: Das auf der GPT-Architektur basierte große multimodale Sprachmodell wurde von OpenAI entwickelt. Im Vergleich zu GPT-3, das ausschließlich natürliche Sprache verarbeiten kann, ist GPT-4 in der Lage, auch Bilder, Dokumente oder Code basierend auf den Eingaben zu generieren, die es erhält. Das Modell wurde mit einer Vielzahl an Daten trainiert und durch ein unüberwachtes Nachtrainieren sowie menschlichen Feedback (RLHF) verbessert.

Die Evaluation erfolgte auf 73 gelabelten Leistungsbeschreibungen, die aus realen Rechnungen extrahiert, gelabelt und durch die WTS bereitgestellt wurden. Zwecks Sicherstellung des Datenschutzes wurden diese entsprechend anonymisiert. Bei den Leistungsbeschreibungen handelt es sich sowohl um Waren- als auch um Dienstleistungen. Insgesamt wurden neun verschiedene Prompts entwickelt – vier Hard-Prompts (Prompt 1-4), drei ICL-basierte Prompts (Prompt 5-7) sowie zwei Few-shot-Prompts (Prompt 8-9).

Die Abfragen wurden so formuliert, dass das Sprachmodell zwei Antworten zurückliefern soll, entweder „Ja“ (d.h. Leistungsbeschreibung berechtigt zum Vorsteuerabzug) oder „Nein“ (d.h. Leistungsbeschreibung berechtigt nicht zum Vorsteuerabzug)⁵. Nachfolgend sind die unterschiedlichen Ansätze sowie deren Abfragen näher erläutert:

- › **Hard-Prompts:** Um den Einfluss der Abfrage auf die Klassifikation zu bewerten, wurden unterschiedliche Formulierungen gewählt. So wurden bspw. Anfragen mit oder ohne Rollenbeschreibung getestet, leicht sprachlich umformuliert oder weitere Hinweise, die nützlich zur Lösung der Aufgabenstellung sein könnten, mitangegeben (z. B.: Ist ein Vorsteuerabzug nach § 14 UStG in Deutschland möglich?).
- › **In-Context Learning:** Für die Umsetzung des ICL-Ansatzes wurde genau eine Abfrage definiert, die mit unterschiedlichen Kontextinformationen angereichert wurde. So wurden bspw. allgemeine oder spezifischere Informationen in der Anfrage mit angegeben, wie eine Leistungsbeschreibung definiert bzw. wann ein Vorsteuerabzug möglich ist.
- › **Few-shot Prompts:** Hier wurde ebenfalls eine Abfrage definiert, die jeweils mit zwei unterschiedlichen Beispiel-Datensätzen angereichert wurde. Bei den Beispiel-Datensätzen handelt es sich zum einen um einen, der sowohl vorsteuerabzugsberechtigte als auch nicht vorsteuerabzugsberechtigte

³ <https://huggingface.co/LeoLM/leo-hessianai-13b-chat>

⁴ <https://openai.com/gpt-4>

⁵ So wurde explizit die Anweisung mitangegeben: „Antworte mit Ja oder Nein. Fall du die Antwort nicht kennst, antworte mit Nein.“

Leistungsbeschreibungen enthielt. Der zweite Beispiel-Datensatz basiert auf einer Blacklist mit Leistungsbeschreibungen, die ausschließlich nicht zum Vorsteuerabzug berechtigen.

Die Evaluation der Ergebnisse erfolgte anhand gängiger ML-Metriken wie Accuracy, Precision, Recall und F1-Score. Diese wurden dabei für jede Klasse berechnet. In erster Linie, um besser verstehen zu können, wie gut das Modell diese vorhersagt.

Ergebnisse und Fazit

Die Evaluation zeigte, dass Prompt-Tuning-Ansätze durchaus eine ressourceneffiziente Möglichkeit zur Feinabstimmung großer Sprachmodelle für die Bewertung von Leistungsbeschreibungen darstellen. Insbesondere beim *GPT-4* zeigte sich, dass die Aufforderungen im Prompt sehr gut verstanden werden. So wurden in allen Prompts die Fragen mit „Ja“ oder „Nein“ beantwortet. Abhängig des gewählten Prompt-Tuning-Ansatzes wurden zudem gute, aber auch schlechtere Ergebnisse erzielt.

Anders im Fall von *LeoLM*, dass neben den forcierten Antwortmöglichkeiten oft eine längere Antwort bzw. oft auch keine Antwort zurücklieferte. In den Antworten von *LeoLM* konnten zudem vereinzelt Halluzinationen bzw. Falschinformationen beobachtet werden. Ein möglicher Grund dafür könnte sein, dass *LeoLM* mit seinen 13 Milliarden Parametern deutlich kleiner als *GPT-4* ist und somit eine wesentlich kleinere Wissensbasis aufweist. Dies erklärt auch die im Schnitt schlechteren Ergebnisse, wie in Abbildung 4 dargestellt.

GPT-4	LeoLM																			
	Accuracy		Precision		Recall		F1-Score		#Predictions		Accuracy		Precision		Recall		F1-Score		#Predictions	
	Ja	Nein	Ja	Nein	Ja	Nein	Ja	Nein	Ja (n=53)	Nein (n=20)	Ja	Nein	Ja	Nein	Ja	Nein	Ja	Nein	Ja (n=53)	Nein (n=20)
Prompt 1	0.69	0.8	0.45	0.77	0.5	0.78	0.47	51	22	0.64	0.78	0.38	0.69	0.5	0.74	0.43	47	26		
Prompt 2	0.75	0.81	0.55	0.84	0.5	0.83	0.52	55	18	0.49	0.67	0.18	0.58	0.25	0.62	0.21	46	27		
Prompt 3	0.69	0.84	0.46	0.71	0.65	0.77	0.54	45	28	0.54	0.7	0.24	0.64	0.3	0.67	0.26	48	25		
Prompt 4	0.49	0.9	0.33	0.33	0.9	0.49	0.49	20	53	0.31	0.63	0.25	0.13	0.8	0.21	0.39	11	62		
Prompt 5	0.73	0.9	0.51	0.71	0.8	0.8	0.62	42	31	0.75	0.75	0.75	0.98	0.15	0.85	0.25	69	4		
Prompt 6	0.79	0.86	0.61	0.84	0.65	0.85	0.63	52	21	0.34	0.72	0.27	0.15	0.85	0.25	0.41	11	62		
Prompt 7	0.75	0.74	1	1	0.1	0.85	0.18	71	2	0.54	0.75	0.3	0.56	0.5	0.64	0.37	40	33		
Prompt 8	0.68	0.75	0.38	0.84	0.25	0.79	0.3	60	13	0.34	0.85	0.28	0.11	0.95	0.2	0.44	7	66		
Prompt 9	0.3	1	0.28	0.03	1	0.07	0.43	2	71	0.32	0.75	0.27	0.11	0.9	0.19	0.42	8	65		

Prompt 1-4: Hard Prompts; Prompt 5-7: In-Context Learning; Prompt 8-9: Few-shot Prompts

Abbildung 4: Evaluationsergebnisse aller Prompts für *GPT-4* und *LeoLM*

Im Rahmen eines Prompt-Tuning sollte man sich jedoch immer bewusst darüber sein, dass ein detaillierter oder zu eingeschränkter Prompt das Modell in eine bestimmte Richtung lenken kann. Eine Richtung, die nicht unbedingt optimal oder sogar irreführend zur Lösung der spezifischen Aufgabe beiträgt. So konnte durch die Angabe einer Rollenbeschreibung (Prompt 2) mit *GPT-4* die Accuracy im Vergleich zu einem einfachen Prompt nur mit der Frage (Prompt 1) verbessert werden (0.75). Durch die Angabe weiterer Anforderungen (z. B. Berücksichtigung § 14 UStG (Prompt 3)) oder durch eine sprachliche Umformulierung (Prompt 4) zeigte sich jedoch, dass sich die Vorhersage deutlich verschlechterte (0.69 bzw. 0.49).

Außerdem kann das Modell seine Fähigkeit verlieren, über den spezifischen Kontext der Aufforderung hinaus zu verallgemeinern, wenn es zu sehr auf die im Prompt angegebenen Beispiele spezialisiert wurde. Dies zeigten die Ergebnisse der Few-shot Prompts (Prompt 8, 9) sowohl mit *GPT-4* als auch *LeoLM*, die deutlich schlechter ausfielen. Insbesondere dann, wenn nur negative Beispiele (Prompt 9) mit angegeben wurden. Hier zeigte sich, dass das Modell sich stark an den Beispielen orientiert und somit deutlich restriktiver und pessimistischer in seiner Klassifikation ist.

Aber auch durch die Berücksichtigung zu allgemeiner Kontextinformationen kann die Genauigkeit der Klassifikation abnehmen. Dies zeigte sich beim ICL. So erzielte zwar *LeoLM* durch die Bereitstellung allgemeiner Kontextinformationen (Prompt 5) eine etwas bessere Accuracy (0.75) als *GPT-4* (0.73), jedoch zeigte sich bei einer detaillierteren Betrachtung der F1-Scores von 0.85 (Ja) und 0.25 (Nein), dass der Prompt sich nur bedingt für den praktischen Einsatz eignen würde. Ein etwas anderes Verhalten zeigte

sich bei *GPT-4*, dessen Accuracy sich durch die Angabe spezifischerer Kontextinformationen (Prompt 6) leicht verbesserte (0.73 auf 0.79).

Zusammenfassend lässt sich sagen: Durch ein Prompt-Tuning können bereits vortrainierte LLMs zur Klassifikation von Leistungsbeschreibungen durchaus zum Einsatz kommen. Für sehr domänenspezifische Fragestellungen, wie sie im Steuerbereich bzw. im Kontext der Beurteilung eines Vorsteuerabzugs vorkommen, implizieren die Ergebnisse jedoch auch, dass ein einfaches Prompt-Tuning nicht ausreichend sein wird. Vor allem wenn es darum geht, den Prozess der Vorprüfung komplett zu automatisieren bzw. eine hohe Genauigkeit bei der Vorhersage sicherzustellen.

Ein Grund dafür ist, dass Leistungsbeschreibungen allgemein sehr vage sein können und oft nur zusammen mit weiteren Informationen (z. B. den Rechnungsdaten) abhängig des Einzelfalls bewertet werden können⁶. Um diese Herausforderung zu berücksichtigen, gibt es jedoch unterschiedliche Verbesserungsmöglichkeiten oder weitere Möglichkeiten, LLMs für spezifische Aufgaben nutzbar zu machen. So stellt bspw. das LLM-Indexing oder Retrieval-Augmented Generation (RAG) Systeme eine weitere Möglichkeit dar, bei dem externe Daten indiziert und so strukturiert, dass das LLM effizient darauf zugreifen kann.. Dadurch lassen sich zusätzliche Informationen aus einer Menge an für die Bewertung der Leistungsbeschreibung relevanter Dokumente berücksichtigen, auf deren Basis schließlich die Beurteilung erfolgt [40]. Weitere Ansätze, um LLM auf die Zielaufgabe effizient anzupassen, werden in der Literatur auch unter dem Begriff Parameter-Efficient Fine-Tuning (PEFT) diskutiert [42].

4.2 Automatisierte Generierung von Verrechnungspreisdokumentationen

Thema & Problemstellung

Zur Bekämpfung künstlicher Verminderungen steuerlicher Bemessungsgrundlagen und grenzüberschreitender Verschiebungen von Unternehmensgewinnen („Base Erosion and Profit Shifting“ bzw. „BEPS“) wurde eine dreistufige Berichtsstruktur geschaffen. Diese setzt sich wie folgt zusammen:

- (1) **Country-by-Country (CbC)-Report:** zur Übersicht aller globalen Aktivitäten
- (2) **Master File:** zur Stammdokumentation
- (3) **Local Files:** zur landesspezifischen Dokumentation

Diese und andere Dokumentationen im Steuerbereich verursachen regelmäßig immense Aufwände. So beziffert bspw. die österreichische Bundesregierung allein die für den CbC-Report pro Unternehmen anfallenden Kosten auf eine Höhe von rund 165.000 Euro. Während für das Master File rund 140.000 Euro und für die erstmalige Erstellung eines Local File mit jeweils 155.000 Euro pro Unternehmen fällig werden [43]. Dies sind Kosten, die je nach Komplexität der Firmenstruktur natürlich auch noch deutlich höher ausfallen können.

Auch wenn regelmäßige Updates nicht so aufwändig sind wie die erstmalige Erstellung, fallen dennoch regelmäßig hohe Kosten an und es braucht entsprechendes Personal, um diesen Dokumentationspflichten nachzukommen. Für multinationale Unternehmen ist es daher unerlässlich, in jedem Land genaue und aktuelle Unterlagen zu führen, in dem diese tätig sind. Sei es nun, um die Verrechnungspreisvorschriften einzuhalten oder ihre Verrechnungspreispraktiken im Falle von Steuerprüfungen bzw. bei Anfragen der Steuerbehörden verteidigen zu können [44].

Dabei erweist sich insbesondere das Local File als besonders aufwändig. Weil es regelmäßigen Änderungen unterliegt und sehr ausführlich alle Transaktionen einer Ländergesellschaft beschreibt. Es enthält detaillierte Informationen über bestimmte konzerninterne Transaktionen, die von der lokalen Einheit in einem bestimmten Land durchgeführt werden. Es enthält Daten über die Art der Transaktionen, die zur Preisermittlung verwendete Methode und alle vergleichbaren Daten oder Benchmarking-Studien,

⁶ BMF v. 01.12.2021 - III C 2 - S 7280-a/19/10002 :001 BStBl 2021/1226715

welche die gewählte Preisermittlungsmethode unterstützen. Die benötigten Informationen werden dabei aus unterschiedlichsten Ursprungssystemen und -dateien zusammengetragen, um schließlich in Textdokumenten mit gewissen Strukturen und Textbausteinen integriert zu werden.

Lösungsarchitektur

Um den zuvor genannten manuellen Aufwand zu minimieren, wurde in Zusammenarbeit mit DFKI, WTS und einer großen deutschen Lebensmitteleinzelhandelsgruppe ein prototypisches KI-Assistenzsystem entwickelt, das Unterstützung bei der Erstellung von Local Files bietet. Als Anwendungsfall wurde dabei die Neuerstellung einer Verrechnungspreisdokumentation für eine neu hinzugekommene Transaktion fokussiert.

Ziel des Assistenzsystems ist hierbei, die relevanten Informationen aus relevanten Eingabedateien zu entnehmen, in ein vordefiniertes Dokumentationsformat anzureichern und entsprechend einzubinden. Insbesondere das Verfassen von Textpassagen in einem professionellen Stil soll dabei abgedeckt werden. In diesem Zusammenhang wurde der Prozess zur Informationsabfrage und insbesondere die Schnittstelle zwischen den einzelnen Ländergesellschaften und der zentralen Steuerabteilung der Dachgesellschaft angepasst bzw. standardisiert.

Schritt 1 – Konzept für ein digitales Formblatt: Konkret wurde ein digitales Formblatt konzipiert, in welchem alle relevanten Informationen von Transaktionsverantwortlichen in den einzelnen Landesgesellschaften eingetragen werden. Hierzu zählen u. a. die stichpunktartige Auflistung der erbrachten Leistungen als Transaktionsbeschreibung, die Funktions- und Risikoanalyse sowie die Auswahl der relevanten Benchmarking-Studien.

Schritt 2 – Erstellung des Local Files: Das erstellte Formblatt dient somit als Ausgangspunkt für den entwickelten Prototypen, welcher nach dem Hochladen des Formblatts ein entsprechendes Local File mit den relevanten Informationen erstellt. Diese Erstellung geschieht zum einen regelbasiert, durch Auswahl relevanter Textbausteine und durch das Einsetzen spezifischer Informationen an der korrekten Stelle. Zum anderen werden mit Hilfe des Luminous-Modells von Aleph Alpha entsprechende Textabschnitte zur Beschreibung der zugrundeliegenden Geschäftsbeziehung sowie der Funktions- und Risikoanalyse generiert.

Schritt 3 – Entwicklung von Prompts: Hierfür wurden diverse Prompts entwickelt, mit denen dem Sprachmodell die Aufgabe nähergebracht und gleichzeitig der notwendige Kontext (wie die eigentliche Transaktionsbeschreibung) mitgeliefert wird. Die Interaktion mit dem Sprachmodell stellt gleichzeitig die Kernkomponente des Prototyps dar.

Schritt 4 – Automatisierte Übersetzung der Inhalte: Eine weitere Funktionalität besteht in der automatisierten Übersetzung der Inhalte mit Hilfe des Sprachmodells, um sowohl die deutsche als auch die englische Variante der Verrechnungspreisdokumentation zu generieren. Dadurch entfällt eine manuelle Übersetzung und/oder die Notwendigkeit eines externen Übersetzungsdiensts.

Schritt 5 – Einbettung des Assistenzsystems: Das Assistenzsystem wurde in einer prototypischen Webanwendung eingebettet, die eine nutzerfreundliche Bedienung des Systems ermöglichen soll und damit den Ablauf zur Erstellung einer Transaktionsbeschreibung maßgeblich optimiert. Innerhalb dieser Webanwendung kann der Endanwender die Eingabedateien hochladen, anhand derer die einzelnen Textbausteine generiert werden. Bei Bedarf lassen sich die generierten Inhalte im Anschluss dann anpassen, z. B. durch die manuelle Korrektur der generierten Formulierungen. Abschließend kann das finale Textdokument heruntergeladen und schlussendlich in die Gesamtdokumentation eingebunden werden.

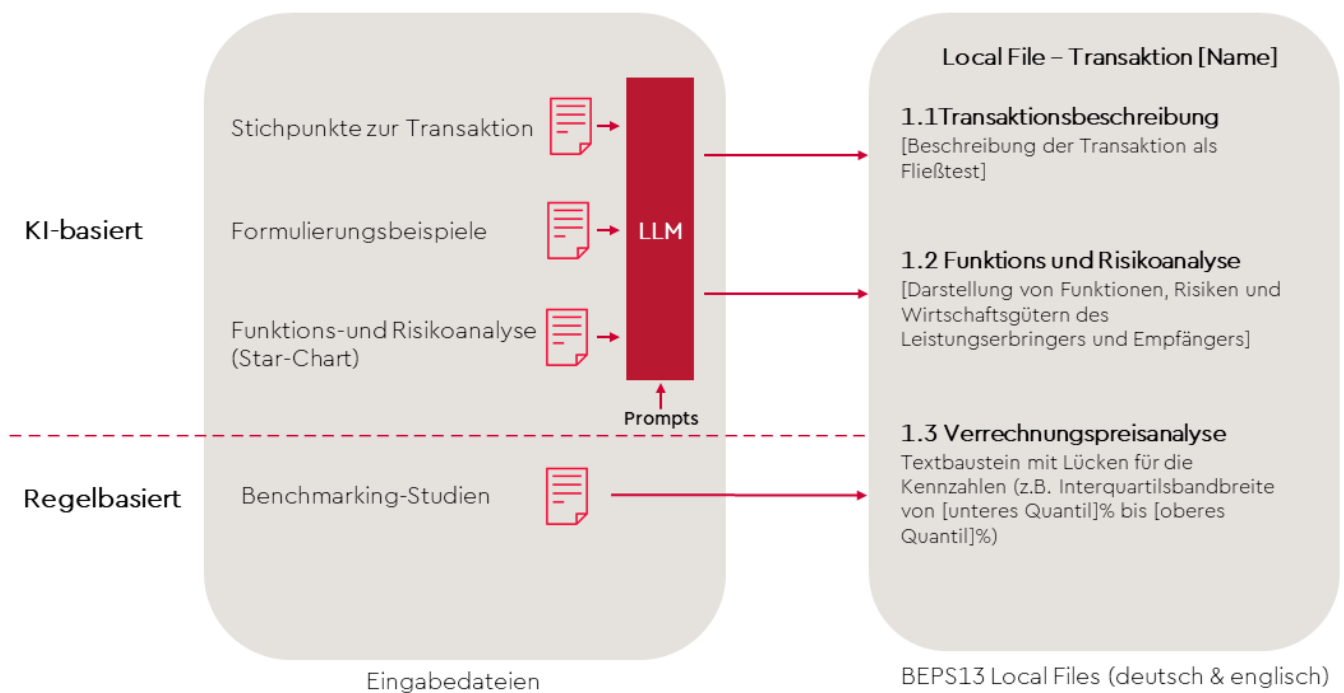


Abbildung 5: Lösungsarchitektur des Assistenzsystems

Ergebnisse und Fazit

Es hat sich gezeigt, dass das Sprachmodell bereits ohne weiterführendes Fine-Tuning für diesen Anwendungsfall geeignet ist. Ebenfalls konnten auch mit einem Zero-Shot-Learning-Ansatz, d.h. dass dem Modell keine Beispiele für produzierende Texte (bspw. die textuellen Darstellungen einer Transaktionsbeschreibung) bereitgestellt werden, bereits gute Ergebnisse erzielt werden. Allerdings muss ein entsprechend hoher Aufwand betrieben werden, um die Prompts so zu gestalten, dass die gewünschten Ergebnisse in zufriedenstellender Qualität produziert werden (Stichpunkt Prompt Engineering).

Grundsätzlich bieten große Sprachmodelle eine wesentliche Hilfestellung bei Aufgaben wie der Texterstellung. Gleichzeitig lassen sich aber auch diverse Herausforderungen feststellen, wie etwa die Anfälligkeit zu halluzinierenden textuellen Darstellungen die sich nur in einem begrenzten Umfang (bspw. Anpassung der Parameter zur Textgenerierung) umgehen lassen. Ebenfalls besteht eine Limitation im beschränkten Weltverständnis der Sprachmodelle bezüglich der Steuerdomäne, sofern kein Fine-Tuning mit steuerbezogenen Daten stattfindet.

Die daraus resultierenden Texte, bspw. Darstellungen im Rahmen der Funktions- und Risikoanalyse, warum beteiligte Gesellschaften bei einer gewissen Transaktion Risiken tragen, sind daher oft eher oberflächlich. Aber auch hier lassen sich Maßnahmen ergreifen, welche diese Beobachtungen reduzieren können: Bei den Entwicklungen wurden bspw. den Prompts die Transaktionsbeschreibungen als zusätzlichen Kontext bereitgestellt und seitens des Modells gefordert, sich auf die jeweiligen Inhalte zu beziehen. Damit konnte ermöglicht werden, dass die resultierenden Texte einen stärkeren Bezug zur jeweiligen Domäne aufweisen.

Insgesamt betrachtet lässt sich festhalten, dass der Einsatz von großen Sprachmodellen für Aufgaben wie die Erstellung der Verrechnungspreisdokumentation als ersten Formulierungsentwurf verstanden werden sollten. Diese Entwürfe sollten anschließend manuell korrigiert und verfeinert werden, um die Sprachqualität, aber vor allem auch inhaltliche Aussagekraft zu verbessern. Nichtsdestotrotz ermöglichen sie eine erhebliche Reduzierung des Aufwands zur Erstellung von Verrechnungspreisdokumentationen.

4.3 Informationsextraktion aus Grundsteuerbescheiden

Thema & Problemstellung

Die Dokumentenverarbeitung stellt einen zentralen Bestandteil der alltäglichen Unternehmenstätigkeiten dar. Insbesondere die Extraktion der in Dokumenten enthaltenen Informationen und deren anschließende Weiterverarbeitung (bspw. Transfer in relevante Informationssysteme) ist ein häufig wiederkehrender Arbeitsablauf. Hinzu kommt, dass papierbasierte Dokumente nach wie vor eine große Rolle beim Informationsaustausch im unternehmerischen Kontext (im Sinne von B2B, B2C und B2G) spielen [45] und damit zentrale Informationen über die unternehmerischen Geschäftsbeziehungen beinhalten. Oft unterliegen diese Informationen zusätzlichen Compliance-Anforderungen, wodurch eine fehlerfreie Erfassung von Informationen unerlässlich ist.

Aufgrund der zuvor geschilderten Gegebenheiten und dem gleichzeitig hohen Automatisierungspotenzial hat sich die Dokumentenverarbeitung als ein prädestinierter Anwendungsfall für RPA entwickelt [14], [46]. Durch die oft analogen Speicherformate liegen die enthaltenen Informationen ohne weitere Verarbeitungsschritte nicht in einem maschinenlesbaren Format vor. Zusätzlich weisen entsprechende Dokumente häufig unstrukturierte oder semistrukturierte Formate mit abweichenden Layouts auf, bedingt durch unterschiedliche Absender der Dokumente. So greifen Unternehmen in der Regel auf ihre eigenen Rechnungsformate zurück und Finanzämter nutzen eigene Vorlagen für die von ihnen ausgestellten Steuerbescheide. Dadurch wird es für automatisierte Extraktionssysteme recht schwer, entsprechende Daten zu identifizieren und auszulesen.

Dank der technologischen Fortschritte im Bereich Deep Learning der letzten Jahre wurde verstärkt erforscht, inwiefern Deep Learning für die Informationsextraktion aus komplexen Dokumententypen realisiert werden kann. Dabei wurden bemerkenswerte Verbesserungen (im Sinne der Extraktionsergebnisse) gegenüber zuvor existierenden regelbasierten und einfacheren ML-Verfahren erzielt. Mit Hilfe dieser Deep Learning Konzepte ist es also möglich, dass sich mit entsprechenden KI-Modellen nicht nur rein textuelle Inhalte (basierend auf OCR), sondern auch visuelle und layout-orientiert Informationen der Dokumente verarbeiten lassen. So können bspw. fettgedruckte Wörter oder Wörter mit erhöhter Schriftgröße etwaige Hinweise auf zentrale Inhalte innerhalb der Dokumente geben und zugleich eine Orientierungshilfe bieten. Solche visuellen Hinweise werden bei einer manuellen Dokumentenverarbeitung durch einen Sachbearbeiter ebenfalls implizit und explizit genutzt.

Die Forschung befasst sich in jüngster Vergangenheit intensiv damit, wie sich diese verschiedenen Informationsmodalitäten bestmöglich in KI-Modelle integrieren lassen. Insgesamt erfolgt somit eine intelligente Informationsextraktion, die zukünftig über eine rein sequenzielle Textverarbeitung weit hinausgeht.

Lösungsarchitektur

In diesem Zusammenhang wurde in Kooperation zwischen WTS und DFKI ein Prototyp entwickelt, welcher eine Deep-Learning-basierte Informationsextraktion aus Einheitswertbescheiden ermöglicht. Hintergrund ist die derzeitige Grundsteuerreform und den dadurch bedingten Verwaltungsaufwänden seitens Steuerabteilungen und steuerpflichtigen Mandanten. Konkret dient die entwickelte KI-Lösung dazu, sämtliche Grundsteuerwert- und Grundsteuermessbetragsbescheide sowie Grundsteuerbescheide der einzelnen Mandanten automatisiert auszulesen und die darin enthaltenen Informationen in einer zentralen Datenbasis aufzubereiten.

Ziel ist es, einen umfassenden Überblick über die Grundbesitzverhältnisse und deren zentralen Kennzahlen wie bspw. Einheitswerte bzw. Grundsteuerwerte, Grundstücksarten und Bebauungen zu erfassen. Basierend auf dieser Datengrundlage kann in nachgelagerten Prüfschritten untersucht werden, ob bspw. die von Finanzämtern bzw. Gemeinden festgesetzten Werte auch den (intern) erwarteten Werten entsprechen.

Durch die Extraktion des Dokumentendatums aus den Grundsteuerwert- und Grundsteuermessbetragsbescheiden kann ebenfalls erfasst werden, bis wann (im Falle von Abweichungen) ein entsprechender Einspruch eingelegt werden kann. Der entwickelte Prototyp setzt an der Informationsextraktion der Einheitswert- bzw. Grundsteuerwert- und Grundsteuermessbetragsbescheide an und liefert somit die Basis für die geschilderten nachgelagerten Prüfschritte. Der Prototyp basiert auf einem hybriden System bestehend aus zwei grundlegenden Komponenten, um die jeweiligen Stärken und Schwächen der beiden Methoden auszunutzen und eine robustere Extraktion zu ermöglichen (vgl. Abbildung 6).

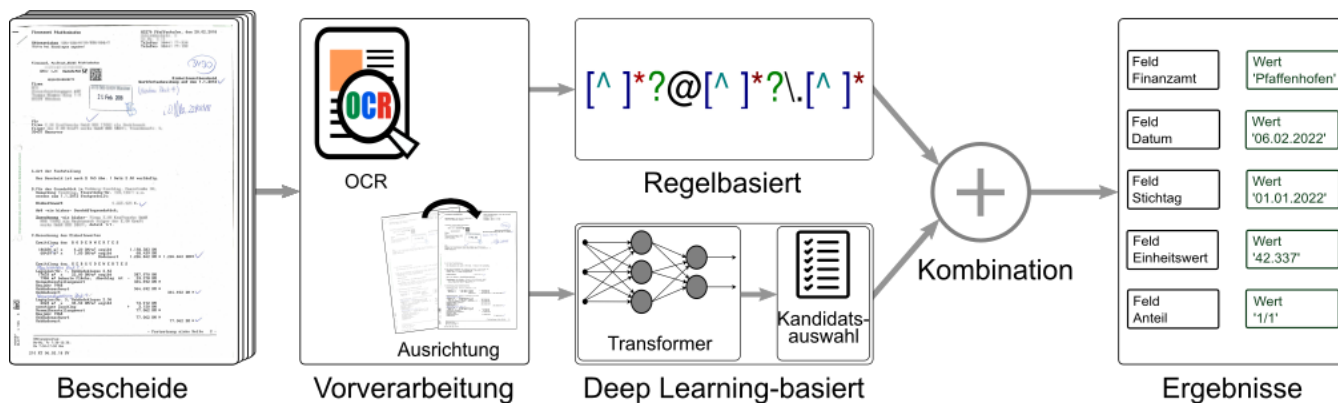


Abbildung 6: Konzeption der Informationsextraktion

Das Extraktionssystem selbst wurde in einer separat entwickelten Webanwendung eingebettet. Ziel dabei ist, die Verwendung so einfach wie möglich zu gestalten – unabhängig vom technischen Hintergrund der jeweiligen Endanwender. Dazu ermöglicht die Webanwendung einen Upload der zu verarbeitenden Dokumente, welche auch zu Aufgabengruppen zusammengefasst werden können, sowie die anschließende Ausführung des Extraktionssystems im Hintergrund.

Schritt 1 – Zusätzliche Vorverarbeitungsschritte: Vor der eigentlichen Extraktion werden verschiedene Vorverarbeitungsschritte durchgeführt, um die Bildqualität und dadurch bedingt die Extraktionsqualität zu verbessern. Ziel dabei ist, bspw. schlechte Dokumentenscans mit schiefer Ausrichtung oder evtl. vorhandene Scan-Artefakte zu reduzieren bzw. zu beseitigen. Bei verschiedenen Tests hat sich herausgestellt, dass insbesondere die korrekte Dokumentenausrichtung maßgeblich zur Ergebnisverbesserung beitragen konnte. Denn dadurch muss das OCR-System keine schiefen Texte verarbeiten, was nicht selten zu Fehlern führt.

Schritt 2 – Regelbasierte Verarbeitung: Der regelbasierte Teil wird durch sog. reguläre Ausdrücke umgesetzt, die auf dem extrahierten Text der Dokumente agieren und innerhalb dieser Texte nach gewissen, vordefinierten Mustern und Formulierungen suchen. Immer mit dem Ziel, die gewünschten Informationen zu extrahieren. Ein solches regelbasiertes Vorgehen äußert sich zwar durch einen relativ hohen Aufwand, der für die Erstellung entsprechender Regeln aufgebracht werden muss, kann aber dennoch sehr gute Extraktionsergebnisse liefern. Zudem bietet es aufgrund der Verständlichkeit zugehöriger Regeln bzw. Muster zugleich ein hohes Maß an Transparenz [47].

Schritt 3 – Deep-Learning-basierte Verarbeitung: Diese Komponente basiert auf der Transformer-Architektur. Da sie ursprünglich für die Verarbeitung linearer Textsequenzen konzipiert wurde, musste sie für den hier vorliegenden Anwendungsfall erweitert werden. Dadurch ist das Modell nun in der Lage, Informationen zu Positionen von Wörtern im Dokument sowie visuelle Informationen der Dokumente in Form von Pixeldaten zu erfassen. Positionsinformationen werden durch Koordinaten sowie Breite und Höhe der Wörter auf den Dokumenten angegeben. Die Integration dieser Informationsquellen ermöglicht dem Modell, positions-orientierte Beziehungen zwischen den Wörtern auf den Dokumenten zu lernen. Die

Informationsextraktion wird dadurch gelöst, dass das Modell für jedes identifizierte Wort auf einem Dokument eine Klasse (bspw. Aktenzeichen) vorhersagt. Basierend auf diesen Vorhersagen können dann für die jeweiligen Klassen entsprechend zugeordnete Wörter erfasst werden.

Schritt 4 - Validierung der Ergebnisse: Sobald die Extraktion abgeschlossen ist, kann der Benutzer die Validierung der Ergebnisse durchführen. Dadurch lässt sich sicherzustellen, dass die Inhalte korrekt extrahiert wurden. Hierfür wurde eine Benutzeroberfläche (vgl.

Abbildung 7) entwickelt, welche einerseits das zugrundeliegende, navigierbare Dokument und andererseits die extrahierten Texte der einzelnen Klassen bzw. Felder übersichtlich und nach Bescheiden getrennt anzeigt. Hier können vom Benutzer die jeweiligen Texte angepasst und/oder ergänzt bzw. entfernt werden, um potenzielle Extraktionsfehler zu entfernen.

The screenshot displays a web application for document validation. The main window shows a scanned document with red boxes highlighting extracted text. The document is titled 'Einheitswertbescheid' and 'Wertfortschreibung auf den 1.1.2012'. The sidebar on the right lists the extracted fields and their values, such as 'Einheitswertbescheid', 'Feststellungsart', and 'Wertfortschreibung'. The interface includes navigation buttons at the top and a sidebar with a list of extracted items.

Abbildung 7: Benutzeroberfläche zur Datenvalidierung

Ergebnisse und Fazit

Die Prototypentwicklung, also die Regelentwicklung wie auch das Modelltraining, wurde anhand von rund 2200 zur Verfügung gestellten Einheitswertbescheiden umgesetzt. Die zugrundeliegende Datenbasis setzt sich aus vielen unterschiedlichen Layouts zusammen und umfasst eine sehr große Zeitspanne mit teilweise über 30 Jahre alten Einheitswertbescheiden. Diese heterogene Datengrundlage ermöglicht es, u. a. Aussagen darüber zu treffen wie generalisierend das System auf unbekannte Formate ist.

Eine quantitative Evaluation, die misst, wie gut die extrahierten Informationen (in textueller Form) gegenüber den tatsächlich korrekten Informationen sind, hat ergeben, dass das System ca. 86 % der Informationen korrekt und fehlerfrei extrahiert. Auffällig ist, dass sich die Ergebnisse je nach zu extrahierendem Feld stärker unterscheiden. Beispielsweise wurde das Aktenzeichen in 97,5 % der Fälle fehlerfrei extrahiert. Auf der anderen Seite wurden die Angaben zu den Straßen der Grundstücke nur zu 45 % fehlerfrei erkannt. Der Hauptgrund für diese Diskrepanzen liegt an der OCR-Extraktion, welche die auf den Dokumenten enthaltenen Texte nicht immer korrekt erkennt.

Aufgrund dessen sind Informationen zu Straßennamen inkl. Hausnummern oder andere Zusätze mit langen zugehörigen Texten sehr fehleranfällig. Und zwar in dem Sinne, dass innerhalb dieser Texte mit hoher Wahrscheinlichkeit mindestens ein falsch erkanntes Zeichen vorliegt. Dadurch können der extrahierte und tatsächliche Wert minimal voneinander abweichen. In diesem Zusammenhang sei aber auch erwähnt, dass solche geringfügigen Extraktionsfehler bei gewissen Feldern wie Straßennamen weniger gravierend sind. Kritische Felder wie das Aktenzeichen, das für eine eindeutige Identifikation der Grundsteuerbescheide notwendig ist, wurden wie oben geschildert sehr gut extrahiert.

Zusätzlich wurde eine erste Umfrage mit dem Personenkreis erhoben, der mit dem Prototyp gearbeitet hat. Das Ergebnis: Durch den Einsatz der KI-Lösung kann die Prozesslaufzeit der Informationsextraktion gegenüber einer rein manuellen Verarbeitung um 25 % bis 50 % reduziert werden. Diese Effizienzsteigerungen resultieren u. a. aus der Tatsache, dass der Endanwender die PDF-Dokumente nicht mehr manuell durchgehen, Bescheide identifizieren und anschließend die enthaltenen Informationen auslesen muss. Außerdem übernimmt der Prototyp die Übertragung der extrahierten Inhalte in ein konsistentes Datenformat. Dadurch müssen die Informationen nicht mehr manuell in eine andere Stelle, bspw. eine Exceldatei, übertragen werden. Folglich wird die Arbeitslast des Sachbearbeiters auf die reine Validierung der automatisiert extrahierten Informationen reduziert.

Aus der Entwicklung des Prototypens ging hervor, dass eine zentrale Herausforderung von automatisierten Extraktionssystemen in der Praxis das zugrundeliegende OCR-System ist. Insbesondere wenn externe OCR-Systeme verwendet werden, die nicht auf den eigenen Dokumenten der Organisation feinjustiert wurden. Denn dann kann es zu vielen falsch erkannten Zeichen kommen, die sich negativ auf das nachgelagerte System zur Informationsextraktion auswirken.

Neben der eigentlichen Performanz des OCR-Systems kann auch die Scanqualität die Extraktionsergebnisse erheblich beeinflussen, wie eingangs bereits geschildert. Insbesondere Stempel, Unterschriften und weitere handschriftliche Notizen kommen nicht selten in der Praxis vor. Allerdings können all diese Elemente die zu extrahierenden Inhalte überdecken und damit die Texterkennung verschlechtern [48]. Gerade für Anwendungsfälle, bei denen die extrahierten Daten ausnahmslos fehlerfrei sein müssen, wird somit die Einbindung eines anschließenden Validierungsprozesses notwendig, im Zuge dessen eine manuelle Validierung der zuvor extrahierten Inhalte stattfindet [46], [48].

4.4 Automatisierung des Transfer Pricing Benchmarking Prozess

Thema & Problemstellung

Verrechnungspreise sind in steuerlicher Hinsicht als „Preise und Konditionen für grenzüberschreitende Geschäftsbeziehungen zwischen verbundenen Unternehmen sowie zwischen Stammhaus und Betriebsstätte“ [47] zu bewerten. Es geht also im Speziellen um die wertschöpfungsadäquate Verteilung des Steuersubstrats aus den an einer Transaktion beteiligten Staaten.

Eine Möglichkeit zum Nachweis und zur Dokumentation der Angemessenheit konzerninterner Verrechnungspreise sind Datenbankstudien. Die Prüfung der Angemessenheit der eigenen Verrechnungspreise erfolgt dabei anhand von GuV-Daten öffentlicher Jahresabschlüsse einer Vergleichsgruppe. Die Vergleichsgruppe soll dabei möglichst aus Unternehmen bestehen, die ausschließlich den betroffenen Unternehmensgegenstand (Produkt oder Dienstleistung) sowie die gleiche Unternehmensfunktion ausüben. Man könnte auch von der Suche nach dem „Einhorn“ sprechen, da die Unternehmen möglichst keine breite Palette von Produkten anbieten oder Funktionen ausüben sollen.

Datenbankanalysen sind ausschließlich bei der Angemessenheitsanalyse von Margen der Routineunternehmen anzuerkennen, wenn die strengen Anforderungen an die Vergleichbarkeit der herangezogenen Vergleichsunternehmen sichergestellt sind. Daher werden in der Praxis oftmals Datenbankanalysen erstellt. In einigen bedeutenden Staaten ist dies bereits gesetzlich verpflichtend. Teilweise wird sogar die jährliche, vollständige Erstellung einer Datenbankanalyse gefordert. In diesem Zusammenhang können für das Screening des Marktes zwei grundlegende, arbeitsintensive und aufeinander aufbauende Schritte unterschieden werden:

Quantitatives Screening: Dabei wird die Suchstrategie in einer Unternehmensdatenbank umgesetzt. In erster Linie, um eine Datenbankanfrage zu stellen, die Unternehmen mit zuvor spezifizierten SIC-/NACE-Codes (Branchencode), Keywords, Unabhängigkeitskriterien, Ländern, Jahren sowie der Umsatzgrößen und Renditekennzahlen („Profit Level Indicator“) selektiert. Darüber hinaus werden ebenso erste Unternehmen eliminiert, zu denen keine Finanzkennzahlen in ausreichendem Maße verfügbar sind oder die auf ein dauerhaftes Verlustgeschäft hindeuten. Üblicherweise wird damit ein Zwischenergebnis von ca. 150 bis 1.000 Unternehmen (je nach Selektionskriterien) gewonnen.

Qualitatives Screening: Im Anschluss an das quantitative Screening erfolgt eine manuelle Feinauswahl der Peergroup. Dabei werden alle (!) zuvor gefunden Unternehmen z. B. via Internetrecherche hinsichtlich Vergleichbarkeit (F&R-Profil, Unternehmensgegenstand, Größe etc.) untersucht. Nicht vergleichbare Unternehmen werden aus der Suche eliminiert: Wenn z. B. angebotene Produkte oder ausgeführte Funktionen abweichen oder keine Internetseite vorhanden ist etc. Danach werden noch einmal die Unternehmen der unteren und oberen 25% der Treffer („Interquartilsbandbreite“) eliminiert, sodass ein Ergebnis mit i.d.R. fünf bis 20 Vergleichsunternehmen entsteht [49].

Durch die Notwendigkeit der regelmäßigen Überprüfung und Neuerstellung der Benchmarking-Studien sowie dem hohen manuellen Aufwand wurde von WTS und dem DFKI ein Prototyp zur Teilautomatisierung des Benchmarking-Prozess entwickelt und zum Proof-of-Concept weiterentwickelt. Durch die vielversprechenden Ergebnisse hinsichtlich des Bearbeitungsaufwands und auch der Standardisierung wird eine Produktentwicklung derzeit angestrebt.

Der Prototyp setzt dabei zwei Strategien um: Erstens werden der Prozess standardisiert und die benötigten Werkzeuge integriert, sodass nicht mehr vielfältige Anwendungen parallel zum Einsatz kommen müssen. Zweitens wird der Benutzer beim qualitativen Screening unterstützt, in dem offensichtlich unpassende Unternehmen automatisiert eliminiert werden. Dadurch sinkt der Umfang der zu prüfenden Unternehmen. Um die Entscheidungen bestmöglich zu unterstützen, werden zu den verbliebenen Unternehmen in einem KI-Cockpit zudem relevante Informationen bereitgestellt – in konzentrierter und übersichtlicher Form. Dabei zeigt sich erneut, dass Benchmark-Studien keine

triviale Aufgabe darstellen. Es werden für jede Studie bzw. für die Findung eines Preises einer Transaktion zwischen 2-5 GB an Daten gespeichert und verarbeitet.

Lösungsarchitektur

Das quantitative Screening wird bereits weitestgehend durch umfangreiche Datenbanken abgedeckt und kann durch Abfragen abgebildet werden, die der jeweiligen Suchstrategie entsprechen. Das anschließende qualitative Screening bietet jedoch umfangreiche Potenziale zum Einsatz von künstlicher Intelligenz. Daher ist die erste Selektion von Unternehmen aus der Datenbank der Einstiegspunkt in den Prototyp.

Diese Selektion umfasst üblicherweise 150 - 1.000 Unternehmen, zu denen die Datenbank Informationen wie Unternehmensname, Webseite, Handelsbeschreibung sowie Finanzkennzahlen und Abhängigkeitskriterien enthält, die zum qualitativen Screening herangezogen werden und anhand derer die Ähnlichkeit zum Referenzunternehmen bewertet werden. Dazu wurde ein Analyse-Dashboard entwickelt, um Entscheidungen mit bestmöglicher Unterstützung aller verfügbarer Informationen treffen zu können.

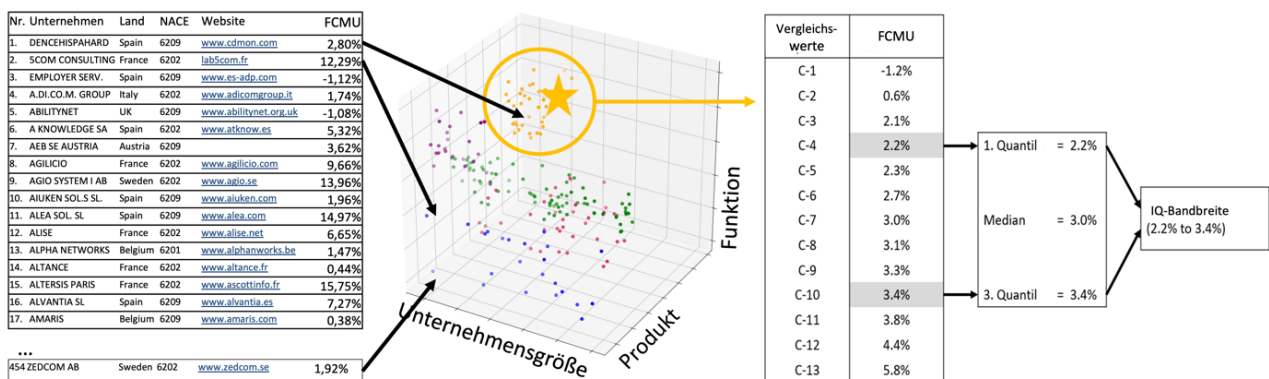


Abbildung 8: Transfer Pricing Benchmarking Prozess

Bei der Analyse kommen gleich mehrere KI-Methoden im Prototyp zur Anwendung:

- (1) **Intelligentes Web Scraping:** Neben verschiedenen Informationen aus den Datenbanken (vor allem Finanzkennzahlen) bieten die Webseiten der einzelnen Unternehmen eine reiche Informationsquelle - sei es über zu untersuchende Kriterien wie die Unternehmensfunktion oder die angebotenen Produkte. Zur Erschließung dieser Informationen durchsucht ein intelligenter Web-Scraper ausgehend der gegebenen Startadresse die gesamte Webseite. Relevante Webseiten werden gespeichert und zur Analyse durch die KI-Systeme weitergeleitet. Relevante Schritte wie die Dokumentation werden dabei automatisch erstellt. Im Zuge dessen eliminieren die KI-Systeme automatisch erste Unternehmen, die stark von Suchkriterien der Nutzereingaben abweichen oder über keine relevanten Informationen verfügen. Die verbleibenden Unternehmen werden dem Benutzer zur Bewertung im Analyse-Dashboard angezeigt. Dort werden die weiteren KI-Methoden der Suchparameter, Klassifikation, Zusammenfassung und Ähnlichkeitsberechnung von Texten übersichtlich dargestellt.
- (2) **Suchparameter:** Suchparameter entsprechen Schlagwörtern, die im Webseitentext des Unternehmens grün/rot hervorgehoben werden können und auf das Akzeptieren oder Eliminieren hinweisen. Der Prototyp schlägt bereits Suchparameter vor, überlässt die Anpassung aber dem Benutzer. Die reine Suche nach Wörtern wurde dabei um eine KI-Komponente ergänzt die zusätzlich auch Synonyme und Wörter mit dem gleichen Wortstamm erkennen kann. So lässt sich bspw. mit dem Suchbegriff „Produkt“ ebenfalls der Begriff „Werkstück“ finden, nicht aber das Wort „Produktion“.

- (3) **Klassifikation hinsichtlich der Unternehmensfunktion:** Bei der Klassifikation werden die Webseitentexte hinsichtlich der Unternehmensfunktion untersucht. Dabei wird geprüft, ob es sich um ein produzierendes, vertreibendes, forschendes, entwickelndes oder verwaltendes Unternehmen handelt oder ob es Dienstleistungen erbringt. Zudem wird in der nächsten Ausbaustufe dem Benutzer die Möglichkeit gegeben, die Vorhersage zu bewerten. Auf diese Weise sollen dem Algorithmus wertvolle Informationen für zukünftige Klassifikationen gegeben werden (Continuous Learning).
- (4) **Textähnlichkeit:** Zu jedem Text wird ein Ähnlichkeitsmaß (0% - 100%) angezeigt. Dieses soll anzeigen, wie ähnlich der Webseitentext zu dem des Referenzunternehmens und zu den Texten der bereits akzeptierten Unternehmen ist. Damit können die potenziell vielversprechendsten Kandidaten für eine bevorzugte Analyse ausgewählt werden. Aber auch stark abweichende Texte lassen sich automatisch eliminieren.
- (5) **Automatische Zusammenfassungen:** Webseiten können teilweise sehr viele Informationen enthalten. So wurden durchschnittlich 1.800 Wörter pro Webseite gezählt. Um die Information zu komprimieren und die Entscheidung zu beschleunigen, wurde eine KI-Anwendung zur Zusammenfassung der Texte implementiert – mit Schwerpunkt auf Unternehmensfunktion, Produkt und Unabhängigkeit des Unternehmens. Dank dieser innovativen Komponente konnte die Wortzahl pro Unternehmen auf 75 reduziert werden. Wodurch sich die Lesedauer von etwa sieben Minuten auf wenige Sekunden reduziert. Dennoch bleibt der Informationsgehalt der gekürzten Texte in 85% der Fälle unverändert, was auch in der Evaluation bestätigt werden konnte.

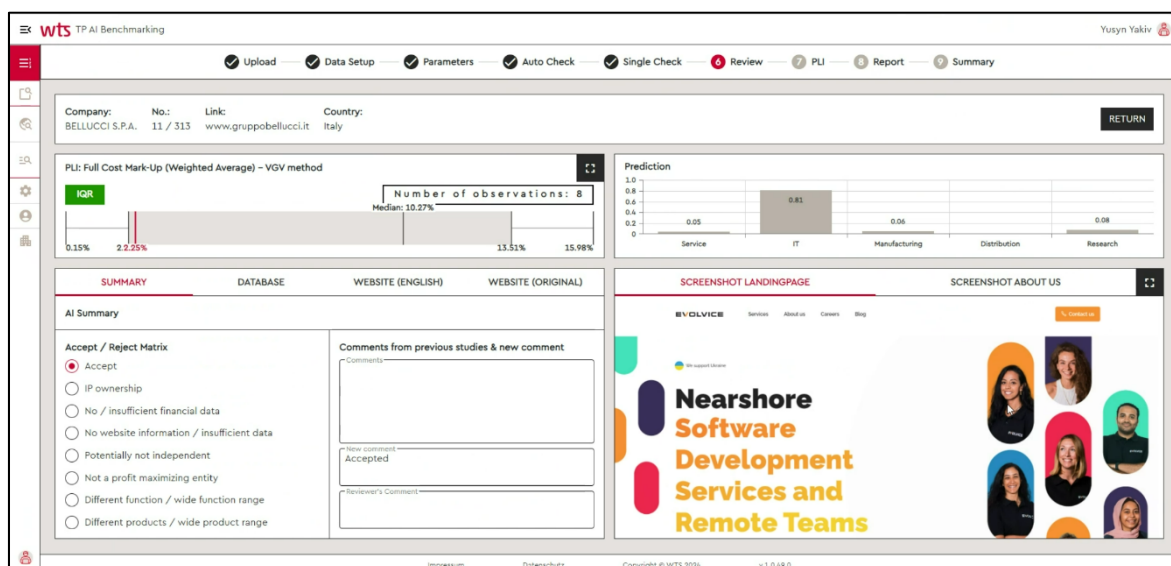


Abbildung 9: User Interface des TP AI-Benchmarking Tools

Ergebnisse und Fazit

Die Evaluation der eingesetzten KI-Methoden erfolgt für jede Methode individuell. Daher werden im Folgenden die Performanz der Textklassifikation, der Textähnlichkeit und der Zusammenfassungen getrennt untersucht.

Textklassifikation: Zur Untersuchung der Methode wurde ein KI-Modell anhand von Unternehmenswebseiten und Labeln von bereits erstellten Benchmarking-Studien trainiert. Dabei wurde ein s.g. LSTM-basierter Ansatz verfolgt, bei dem nicht nur das Auftreten von Wörtern eine Rolle in der Klassifikation spielt. Auch die Reihenfolge der Wörter sowie deren Bedeutung in dem jeweiligen Kontext wurden berücksichtigt [50]. Wegen einer stark unterschiedlichen Anzahl von Texten pro Klasse wurden die Unterschiede durch die Generierung von synthetischen Daten ausgeglichen. Dabei wurden Daten einer Klasse kopiert und die Kopie durch Löschen, Einfügen und Ersetzen von Wörtern variiert. Dadurch konnte bei der Analyse der Unternehmensfunktion bei neuen, bisher unbekanntem Unternehmen eine Vorhersagegenauigkeit von 92% erreicht werden.

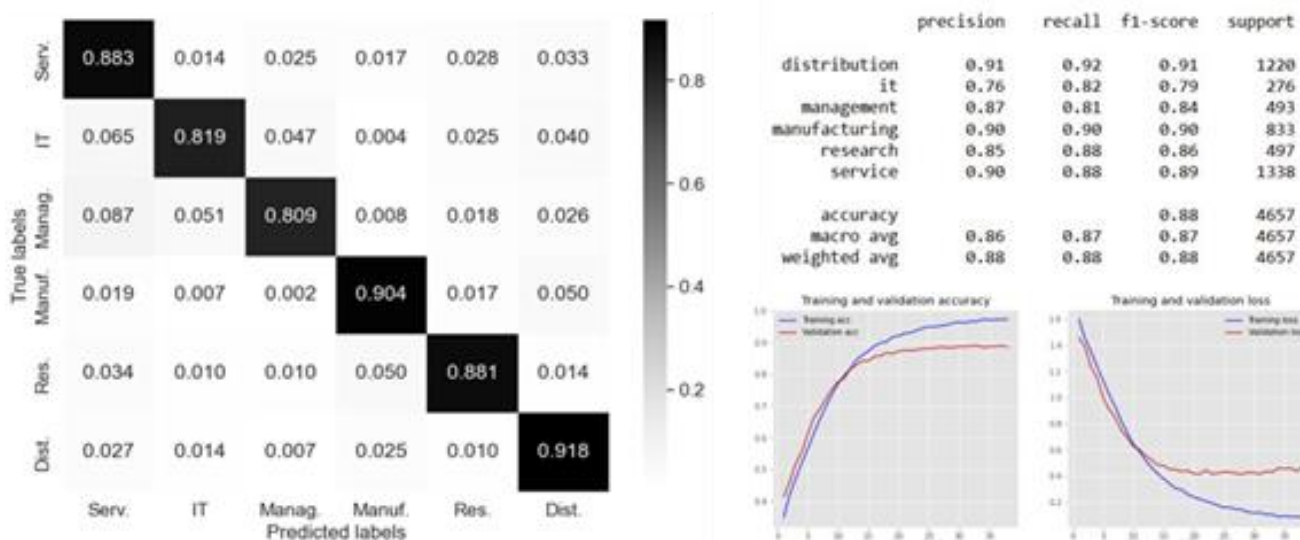


Abbildung 10: Kennzahlen zur Performance der Textklassifikation

Die Kennzahlen zur Performance der Textklassifikation können in Abbildung 10 eingesehen werden. Zu jeder Klassifikation wird auch deren Sicherheit berechnet, sodass diese Analyse bei der automatischen Elimination zum Einsatz kommt. Um Unternehmen mit einer anderen als der gesuchten Funktion sowie mit hoher Sicherheit zu eliminieren. Aufgrund der Komplexität einzelner Unternehmen und der Vielfalt deren Tätigkeiten ist eine Klassifizierung zu einer Klasse nicht immer möglich. Eine niedrige Vorhersagegenauigkeit ist oftmals ein Indikator für Unternehmen, die mehrere Funktionen ausüben.

Textähnlichkeit: Im zweiten Schritt wurde die Ähnlichkeiten von Webseitentexten evaluiert. Der Ansatz fußt auf dem vortrainierten, transformer-basierten Sprachmodell BERT, welches das „Sprachverständnis“ enthält, sowie der Kosinus-Ähnlichkeit als Ähnlichkeitsmaß [51]. Zur Evaluation wurden bestehende Benchmark-Studien herangezogen, die bekannte akzeptierte und eliminierte Unternehmen sowie deren Webseitentexte enthalten. Die Ähnlichkeiten zwischen allen Texten der akzeptierten Unternehmen sowie die Texte aller zu untersuchenden Unternehmen wurde gemessen. Dabei zeigte sich, dass sich die Gruppe der Texte von akzeptierten Unternehmen wesentlich ähnlicher ist als die Texte der akzeptierten und eliminierten Unternehmen zusammen. Konkret betrug das Ähnlichkeitsmaß im Schnitt 73,9% in der Gruppe der akzeptierten Unternehmen gegenüber 49,8% in der gesamten Studie. Die Untersuchung umfasste insgesamt 5 Benchmark-Studien. Die Ergebnisse wurden gemittelt.

Textzusammenfassung: Die dritte Evaluation betrifft die automatische Zusammenfassung von Texten. Die textuelle Zusammenfassung wird dabei mit einem großen Sprachmodell (vgl. Kapitel 2.3) durchgeführt. Wobei durch Prompt-Engineering der Schwerpunkt auf Produkte/Dienstleistungen und Unternehmensfunktion gelegt wird. In der Evaluation haben 3 Probanden eine Stichprobe von 232 Unternehmen bewertet. Dazu wurde jedem Probanden, zu jedem Unternehmen, einmal der vollständige Text der Unternehmenswebseite sowie die automatisch erzeugte Zusammenfassung vorgelegt. Dieser untersuchte

u. a. den Informationsgehalt und die Nützlichkeit der Zusammenfassung für den Business Case im Vergleich zur Webseite und trug seine Ergebnisse in einen Feedbackbogen ein. Der Informationsgehalt konnte auf einer Skala von -5 bis +5 gewählt werden. -5 bedeutet in diesem Fall maximaler Informationsverlust in der Zusammenfassung, 0 einen identischen Informationsgehalt und +5 Information wurden gefunden, die nicht im Text sind. Z. B. durch Halluzination durch das großen Sprachmodells oder auch eine Fehlinterpretation des Probanden. Optimal wäre somit ein Ergebnis von 0. Über alle Bewertungen hinweg wurde ein Mittelwert von -0.16 erreicht, sodass die Zusammenfassungen einen minimalen Informationsverlust erleiden, aber natürlich bei weitem weniger Text enthalten. Die Nützlichkeit war in 78% aller Fälle gegeben, sodass die Zusammenfassung zu einer Entscheidung herangezogen werden konnte. Von den 52 Fällen, in denen es nicht möglich war mit der Zusammenfassung eine Entscheidung zu treffen, war dies in 37 Fällen auch mit der Webseite nicht zweifelsfrei möglich. Die Effektivität beträgt somit 85%.

4.5 Synthetische Daten für die Anomalieerkennung

Thema und Problemstellung

Das Anwendungsszenario befasst sich mit der Anomalieerkennung in Finanzbuchhaltungsdaten. Im Kern handelt es sich dabei um ein Multiklassen-Klassifikationsproblem, da neben regulären Finanztransaktionen zwischen zwei Anomalieklassen unterschieden wird.

Da Klassifikationsalgorithmen im Kontext der Anomalieerkennung auf die Beispiele der Mehrheitsklasse ausgerichtet sind, ist die Minderheitsklasse (Klasse der Anomalien) entsprechend im ML-Modell unterrepräsentiert. Durch die Voreingenommenheit des Modells gegenüber den dominanten Klassen ist es möglich, dass das Modell entsprechend diese bevorzugt. Dadurch wird eine sehr hohe Genauigkeit in der Erkennung dieser erzielt. Demgegenüber wird für seltene, potenziell signifikante Ereignisse jedoch eine deutlich schlechtere Genauigkeit erreicht.

Diese Unausgeglichenheit der Klassenverteilung kann somit zu Problemen bei der Erkennung von Anomalien führen. Entsprechend irreführend können traditionelle Bewertungsmetriken sein, wodurch selbst bei einer Modell-Genauigkeit von 95% offensichtliche Anomalien übersehen werden. Zudem besteht die Gefahr einer schlechten Generalisierung des Modells, falls dieses auf eine entlang der Minderheitsklasse ausgerichteten Stichprobenziehung trainiert wurde. Basieren die getroffenen Entscheidungen des Modells auf falschen Korrelationen, hat dies nicht nur Auswirkung auf die Wirtschaftlichkeit der Prüfung (siehe Falsch-Positiv-Rate und Falsch-Negativ-Rate), sondern auch auf die Interpretierbarkeit des Modells selbst. Denn insbesondere im Finanz- und Steuerbereich ist die Nachvollziehbarkeit von Modellentscheidungen sowohl aus organisatorischen als auch rechtlichen Gründen von großer Bedeutung.

Das Ziel des Anwendungsfalls ist es, aussagekräftige Transaktionen der Minderheitsklasse zu reproduzieren. Dementsprechend geht es im Anwendungsfall darum, selten vorkommende bzw. anomale Transaktionen zu approximieren. Durch die so erreichte Erhöhung der Menge und Variabilität der wenig vorhandenen Daten soll die Genauigkeit und Performanz eines ML-Modells verbessert werden. Für die Generierung synthetischer Anomalien wurden unterschiedliche Ansätze implementiert. Die Evaluation erfolgte anhand gängiger ML-Techniken zur Anomalieerkennung.

Lösungsarchitektur

Der verwendete Datensatz stammt von Schreyer et al. und orientiert sich an der SAP-Buchungsstruktur [52]. Dieser enthält insgesamt sieben kategorische und zwei numerische Attribute, die in den SAP FICO-Tabellen BKPF (mit den gebuchten Buchungsköpfen) und BSEG (mit den gebuchten Buchungssegmenten) verfügbar sind. Der Datensatz umfasst insgesamt 533.009 Finanztransaktionen, wobei insgesamt 100 Transaktionen als anomal gekennzeichnet sind. Diese lassen sich selbst in 30 lokale Anomalien und 70 globale Anomalien unterscheiden. Die Durchführung des Experiments selbst umfasst insgesamt vier Schritte (vgl. Abbildung 11).

Lokale Anomalien: Beschreiben Transaktionen, die eine ungewöhnliche oder seltene Kombination von Attributwerten aufweisen, während die einzelnen Attributwerte recht häufig auftreten, z. B. ungewöhnliche Buchungssätze.

Globale Anomalien: Beschreiben Transaktionen, die ungewöhnliche oder seltene individuelle Attributwerte aufweisen. Diese Anomalien beziehen sich in der Regel auf stark verzerrte Attribute, z. B. selten buchende Benutzer, selten verwendete Konten, sehr hohe Buchungsbeträge oder ungewöhnliche Buchungszeiten.

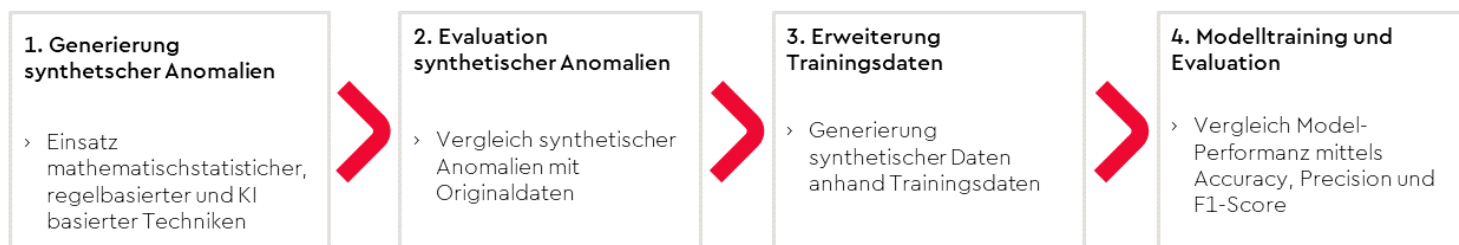


Abbildung 11: Vorgehensweise der Generierung synthetischer Anomalien

Für die Generierung synthetischer Anomalien wurden Techniken einzeln, aber auch in Kombination verwendet. So wurden Techniken eingesetzt, die die Klassenverteilung in einem Datensatz ausgleichen. Und zwar indem Datenpunkte der Minderheitsklasse entsprechend der Mehrheitsklasse zufällig vervielfältigt (Random Oversampling) bzw. zusätzlich die Anzahl der Datenpunkte der Mehrheitsklasse durch Stichprobenziehung reduziert (Over- und Undersampling) wurden. Außerdem wurden neue Datenpunkte entsprechend der relativen Häufigkeit aller kategorischer Attributausprägungen zufällig erzeugt.

Neben rein statistischen Verfahren wurden auch zusätzlich KI-basierte Ansätze verprobt. So wurde sowohl mit einem Vanilla GAN⁷ als auch einem CTGAN⁸, ein auf tabellarische Daten spezialisiertes GAN, neue anomale Datenpunkte generiert. Das Vanilla GAN bzw. CTGAN wurde dazu auf den anomalen Datenpunkten trainiert. Schließlich wurden auch die Möglichkeiten von ChatGPT⁹ näher analysiert, um auf „Knopfdruck“ Code für die Generierung synthetischer Anomalien zu erzeugen.

Die Bewertung der Qualität der Daten erfolgte durch gängige statistische Ansätze wie den Vergleich der Attributverteilungen, Hypothesentests und Korrelationsanalysen. Darauf aufbauend wurden Techniken ausgewählt, welche die Ursprungsverteilung bestmöglich widerspiegeln, und entlang dedizierter ML-Verfahren zur Anomalieerkennung evaluiert.

Um *Data-Leakage* zu vermeiden, wurden ausschließlich auf den im zuvor erstellten Trainingsdatensatz enthaltene Anomalien neue synthetische Anomalien erzeugt und in der Trainingsphase berücksichtigt. Für die Evaluation der Ergebnisse wurden klassische Metriken wie *Accuracy*, *Precision* und *F1-Score* verwendet.

⁷ <https://github.com/NextBrain-ai/nbsynthetic>

⁸ <https://github.com/sdv-dev/SDV>

⁹ <https://chat.openai.com/>

Ergebnisse und Fazit

Synthetische Daten sind eine vielversprechende Technologie, um typische Herausforderungen entlang der Themen wie Modelltraining und -verbesserung oder Datenschutz zu lösen. Trotz des großen Potenzials birgt die Technologie dennoch auch Risiken, die es zu berücksichtigen gilt. Welche Vor- und Nachteile der Einsatz der Technologie hat, wird nachfolgend anhand der erzielten Ergebnisse des Experiments erläutert.

Anhand des Anwendungsfalls konnte gezeigt werden, dass synthetische Daten die Entwicklung beschleunigen und somit Kosten senken können. Insbesondere im Hinblick auf die Bereitstellung entsprechender Daten für ein Modelltraining. Vor allem generative KI-Lösungen wie ChatGPT bergen großes Potenzial, um die Datenstruktur selbstständig erlernen und synthetische Daten generieren zu können. Beispielsweise betrug die absolute durchschnittliche Abweichung der Attributsausprägungen je nach Implementierung 0.25 bzw. 0.05. Wobei ein Wert von 0 eine exakte Kopie der Originaldaten bedeuten würde. Auch das CTGAN erzielte einen guten Wert von 0.25, im Gegensatz zum Vanilla GAN mit 0.87. Jedoch sei hier angemerkt, dass Standard-Parametereinstellungen verwendet wurden und durch eine Optimierung der Hyperparameter mit einer Verbesserung zu rechnen sei.

Grundsätzlich ist an dieser Stelle zu erwähnen, dass der Einsatz entsprechender Techniken auch ein optimales Verständnis über diese erforderlich macht. Dies ist der Tatsache geschuldet, dass es sich um Black-Box-Modelle handelt, die sehr undurchsichtig und komplex sein können. Letztendlich hängt die Qualität stark vom Training des generativen Modells ab, wodurch auch die Qualität der Datenpunkte erheblich variieren kann.

Ferner ist zu beachten, dass synthetische Daten kein Ersatz für reale Daten sind. Je nachdem welcher Ansatz für die Generierung synthetischer Daten gewählt wurde, handelt es sich zwangsläufig immer um eine verzerrte Version der realen Daten. Während synthetische Daten verwendet werden können, um die Performanz bei unausgeglichenen Daten zu verbessern, sollte die Bewertung der Daten immer anhand von realen Daten erfolgen. Beispielsweise lieferte der Ansatz basierend auf relativen Häufigkeiten eine absolute mittlere Abweichung von 0. Das liegt daran, dass die Attributsausprägungen entsprechend der Häufigkeitsverteilung der Originaldaten entsprechen. Anders sah es beispielsweise bei SMOTE (0.59) bzw. K-Means SMOTE (0.29) aus, die eine deutliche Verzerrung aufweisen. Dies liegt u. a. daran, dass die Grenzpunkte zwischen den einzelnen Klassen nicht gut definiert sind (insbesondere bei den lokalen Anomalien). So liegen bspw. die numerischen Datenausprägungen *DMBTR* und *WRBTR* (d.h. die isolierten Punkte) weit in der Region der Mehrheitsklasse. Dadurch kommt es zu einer Überlappung der Grenzpunkte zwischen der Minderheits- und Mehrheitsklasse.

Außerdem sollte im Umgang mit synthetischen Daten darauf geachtet werden, dass keine Informationen außerhalb des Trainingsdatensatzes berücksichtigt werden. Dies würde zu einer übermäßig optimistischen Leistungsschätzung führen. Weil das ML-Modell einen entsprechenden Zugang zu Informationen hat, die diesem in der realen Welt eigentlich nicht zur Verfügung gestanden hätten. Konkret bedeutet das, dass synthetische Daten ausschließlich auf Basis des Trainingsdatensatzes generiert werden sollten. Andernfalls würde die Möglichkeit bestehen, dass im Testdatensatz die gleichen synthetischen Datenpunkte enthalten wären, die auch im Trainingsdatensatz vorhanden sind.

Mit Blick auf die Leistungsverbesserungen wurden gängige Techniken zur Anomalieerkennung wie Random Forest, Isolation Forest, XGBoost und ein Deep Autoencoder implementiert. Während bei Random Forests jeweils eine deutliche Leistungsverbesserung im Vergleich zu den Originaldaten erzielt werden konnte, führten die synthetischen Anomalien beim Isolation Forest jedoch zu keiner wesentlichen bzw. teilweisen Verschlechterung der Leistung. Eine Begründung liegt hier insbesondere in der Funktionsweise und Parametrisierung des Algorithmus, wie anomale Datenpunkte identifiziert werden (d.h. je mehr Splits für die Isolation jedes einzelnen Datenpunktes benötigt werden, desto unwahrscheinlicher handelt es sich um eine Anomalie).

Das mit Abstand beste Ergebnis erzielte der Deep Autoencoder auf den Originaldaten, der in Abhängigkeit des gewählten Schwellwertes für den Rekonstruktionsfehler alle anomalen Transaktionen identifizieren konnte. Eine Erweiterung des Datensatzes mit synthetischen Anomalien war entsprechend nicht mehr erforderlich. Zusammenfassend lässt sich festhalten, dass durch synthetische Daten eine Leistungsverbesserung erzielt werden kann. Entscheidend ist jedoch zum einen die Qualität der Daten, zum anderen die eingesetzte Technik unter Berücksichtigung der Funktionsweise des verwendeten Algorithmus. Auch wenn der Deep Autoencoder sehr gute Ergebnisse erzielt hat, ist das wohl auch der zugrundeliegenden Datenstruktur der anomalen Datenpunkte geschuldet.

Synthetische Daten bietet nicht nur im Hinblick auf eine etwaige Leistungsverbesserung neue Möglichkeiten, sondern auch bezogen auf den Datenschutz und die Privatsphäre. Da keine sensiblen Informationen verwendet werden, bieten synthetische Daten einen erheblichen Schutz vor Verstößen oder Missbrauch. Dadurch eignen sich diese insbesondere in kollaborativen Projekten, in denen eine gemeinsame Nutzung und Analyse erforderlich ist. Zudem lassen sich kollaborativen Umgebungen einfach nutzen, um auf ML-spezifische Hard- und Software zuzugreifen. Allerdings ist der Prozess der Generierung synthetischer Daten mit einigen Herausforderungen verbunden. So muss dieser robust und sicher konzipiert sein. Andernfalls besteht die Möglichkeit, dass synthetische Daten unbeabsichtigt bestimmte Muster des ursprünglichen Datensatzes widerspiegeln. Dadurch besteht wiederum die Gefahr, dass unbeaufsichtigt sensible Details offengelegt werden. Entsprechend erforderlich ist ein kontinuierliches Monitoring der implementierten Verfahren und Techniken.

4.6 Kombination großer Sprachmodelle mit Wissensgraphen

Thema und Problemstellung

Wie bereits in den vorherigen Abschnitten dargelegt, stellt die Extraktion von Informationen aus steuerlichen Dokumenten einen Anwendungsfall für die Nutzung von generativer KI in der Steuerfunktion dar. Allerdings sind bei der Anwendung großer Sprachmodelle im Allgemeinen und insbesondere auch im Steuerbereich große Herausforderungen verbunden:

Es ist sicherzustellen, dass das Modell stets auf Grundlage der aktuellen Gesetzgebung antwortet. Hierfür müssen die KI-Modelle fortlaufend mit neuen Daten angereichert und veraltete Quellen aus der Datengrundlage entfernt werden [51-54]. Des Weiteren besteht die Gefahr, dass KI-Modelle aufgrund fehlender Daten falsche Informationen ausgeben (siehe Halluzination). Ein weiteres Problem können zudem viel zu allgemeine Antworten auf fachspezifische Fragen darstellen. Nicht zu vergessen ist die Gewährleistung einer hinreichenden Nachvollziehbarkeit der KI-Modelle, damit diese in der täglichen Arbeit der Steuerfunktion möglichst effektiv zum Einsatz kommen können.

Um diese Probleme zu vermeiden, sind in der Forschung seit Jahrzehnten Ansätze zur expliziten Wissensrepräsentation entwickelt worden. Wissensgraphen, Entscheidungsbäume, Datenbanken und andere Formen der expliziten Wissensrepräsentation sind insbesondere auch bereits vielfältig im Steuerbereich zum Einsatz gekommen (siehe beispielsweise die DFKI-WTS-KI-Innovationsstudie aus dem Jahr 2017). Eine besondere Herausforderung besteht gerade darin, diese Ansätze mit großen Sprachmodellen zu kombinieren, also hybride Systeme zu entwickeln. In einem hybriden System werden symbolische und sub-symbolische Verfahren zur Wissensrepräsentation kombiniert. Man spricht auch von sogenannten neuro-expliziten Methoden.

Ein möglicher Weg zur Entwicklung hybrider Systeme ist es, dem Modell zunächst Hierarchien und Beziehungen innerhalb der zugrundeliegenden Daten zu vermitteln. Zum Beispiel muss ein KI-Modell bei der Beurteilung steuerlicher Sachverhalte stets dazu in der Lage sein, bestehende Regelungshierarchien konsequent einzuhalten. Dementsprechend muss es lernen, Sachverhalte zuerst auf Basis von Gesetzen zu prüfen. Anschließend gilt es, die gesamten Inhalte mit nachrangigen Quellen wie bspw. aktuellen Urteilen oder Stellungnahmen erneut zu beleuchten.

Lösungsarchitektur

Einigen der oben genannten Herausforderungen kann mit einem sogenannten Retrieval-Augmented-Generation-Ansatz begegnet werden[55], [56]. Dabei wird zur Verbesserung der Ausgabe eines KI-Modells eine weitere externe Datenbasis zur Verfügung gestellt, zusätzlich zu den bereits vorhandenen Trainingsdaten. Auf diese Weise wird das allgemeine Wissen des Modells um fachspezifisches Know-how ergänzt.

Dadurch ist es großen Sprachmodelle möglich, noch vor der Generierung einer Ausgabe auf eben diese externe Ressource zuzugreifen und eine qualitativ hochwertigere Antwort zu liefern. Gegenüber der alleinigen Verwendung eines großen Sprachmodells bietet diese Methode also bereits einige Vorteile. Vor allem im Hinblick auf die Tatsache, dass die Verwendung einer externen Datenquelle den Wissensstand der KI-Modelle kontinuierlich und vor allem kostengünstig mit Fachwissen erweitert – und dieses Wissen auch fortlaufend auf dem neuesten Stand hält.

Die Mehrheit moderner Chatbots greift über eine API auf vortrainierte Sprachmodelle zurück, deren Wissen implizit in deren angelernten Parametern verankert ist. Eine direkte Erweiterung dieser Grundlagen durch aktuelles fach- und unternehmensspezifisches Wissen mit Hilfe eines erneuten Trainings ist zumeist sehr aufwendig und kostenintensiv. Die Verwendung einer externen Datenquelle bietet hier eine praktikable Alternative und bietet Entwicklern eine zusätzliche Kontrolle darüber, welche Daten in das Modell gelangen. Dies hat wiederum zur Folge, dass das Vertrauen in die Sprachmodelle innerhalb einer Organisation gesteigert werden kann[56].

Aus diesen Gründen liefert die Nutzung eines RAG-Ansatzes in Verbindung mit einem Wissensgraph (engl. Knowledge Graphs (KG)) als externe Datenquelle eine vielversprechende Lösung[57]. Denn dieser erweiterte Ansatz vereint nicht nur die bisher genannten Vorteile. Zusätzlich begegnet er auch der Herausforderung, die jeweiligen Dokumentenhierarchien und Beziehungen zu vermitteln und die Nachvollziehbarkeit der Modelle zu verbessern. Ein Wissensgraph ist eine Datenstruktur, mit der sich Informationen in graphischer Form abbilden und verwandte Daten durch verbundene Entitäten repräsentieren lassen[58]. Die Informationen werden dabei als Triple gespeichert: *head entity*, *relation* und *tail entity*. Ein einfaches Beispiel für solch einen Wissensgraphen ist in Abbildung 12 gegeben.

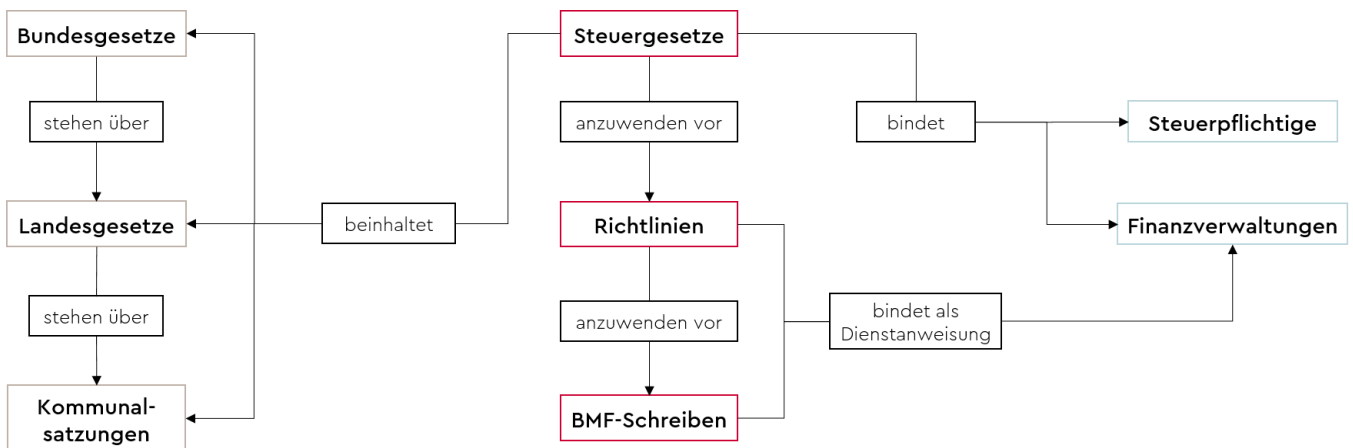


Abbildung 12: Beispiel eines Wissensgraphen für den steuerrechtlichen Rahmen.

Eine Untergruppe von allgemeinen KGs bilden die sogenannten „Domain-specific Knowledge Graphs“, welche es ermöglichen, domänenspezifisches Wissen wiederzugeben [59], [60]. Im Vergleich zu allgemeinen Wissensgraphen sind diese zwar kleiner, liefern aber verlässlichere und exaktere Informationen. KGs können fortlaufend erweitert und aktualisiert werden. Zudem macht es ihr einfacher logischer Aufbau möglich, wertvolle Schlussfolgerungen zu ziehen und interpretierbare Ergebnisse zu liefern[61].

Indem sich Informationen innerhalb eines KGs ins Verhältnis setzen lassen, eröffnet sich ein weiterer Lösungsansatz für die bereits zuvor erwähnte Problematik der Berücksichtigung von Dokumentenhierarchien, was die Beurteilung von steuerlichen Sachverhalten durch generative Sprachmodelle betrifft. All diese Punkte machen es naheliegend, bei Verwendung eines RAG-Ansatzes zur Implementierung eines Steuer-Chatbots auf einen „Domain-specific Knowledge Graph“ als externe Datenquelle zurückzugreifen.

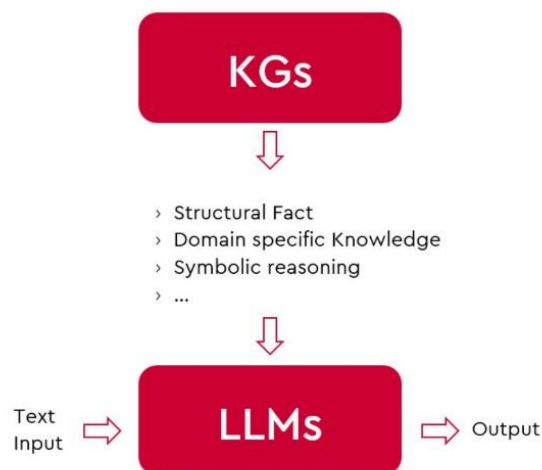


Abbildung 13: Kombination eines großen Sprachmodells mit einem Knowledge Graph [61]

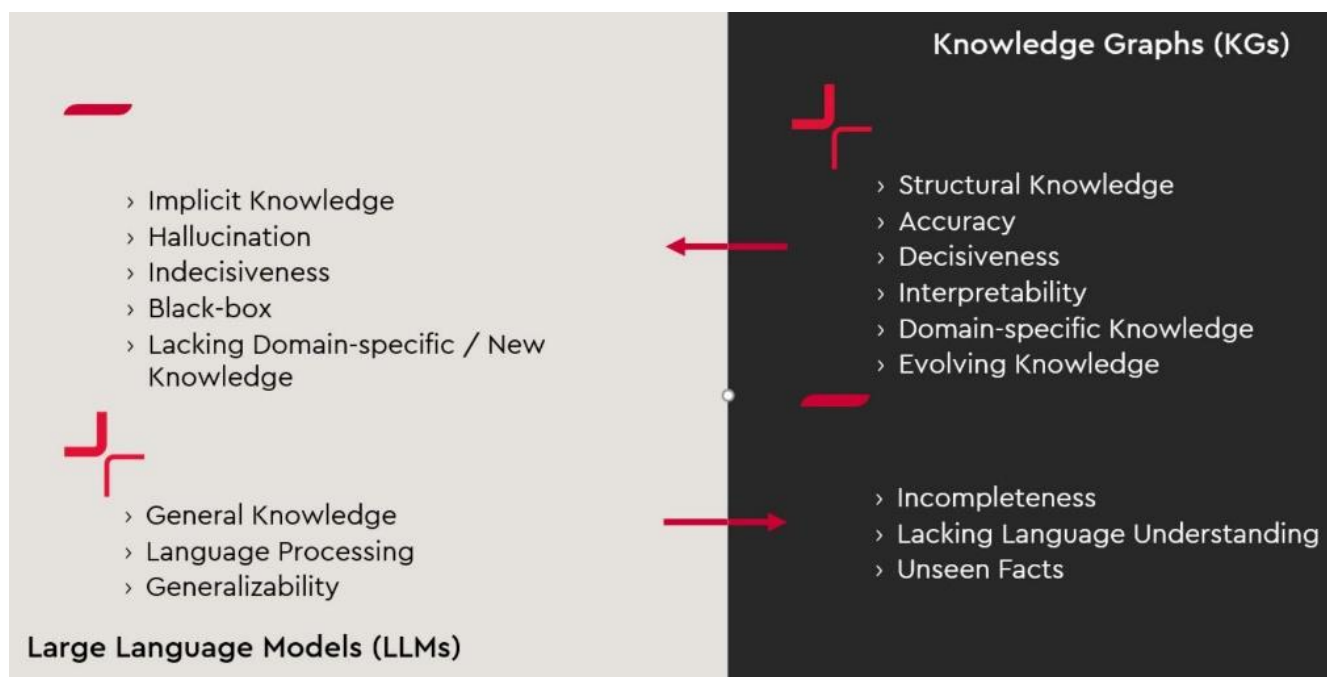


Abbildung 14: Vor- und Nachteile von großen Sprachmodellen und Knowledge Graphs [61]

Die Verbindung aus großen Sprachmodellen und KG komplementiert dabei die Schwächen des jeweils anderen Modells, wodurch wiederum ein leistungsfähiges Gesamtmodell entsteht. Im Vergleich zu Sprachmodellen fehlt es KG an der Fähigkeit, natürliche Sprache zu verarbeiten und auf ein großes Allgemeinwissen zurückzugreifen. Demgegenüber schneiden Wissensgraphen gegenüber Sprachmodellen besser bei der akkuraten Wiedergabe von spezifischem, strukturiertem Wissen sowie bei der Interpretierbarkeit ab[58].

Die Nutzung eines KG ist im Vergleich zu normalen Retrieval-Techniken mit einem zusätzlichen Mehraufwand verbunden. Darüber hinaus sind tiefere Kenntnisse über den eigentlichen Inhalt und die

Struktur erforderlich. Aus diesem Grund ist eine enge Zusammenarbeit zwischen Entwicklungsteams und steuerlicher Expertise nicht nur wünschenswert - vielmehr bildet diese die nötige Basis für eine Implementierung von KG, um das volle Potenzial dieses Ansatzes ausschöpfen zu können.

Ergebnisse und Fazit

Klassische RAG-Ansätze gehören schon heute fest zum Fundament sprachmodell-basierter wissens-basierter Systeme. Dabei zeigt sich deutlich, dass die Zugabe weiterer Fachinformationen die Genauigkeit und Qualität der Antworten deutlich steigern kann. Das trifft insbesondere auf Informationen zu, die aus zeitlicher oder fachspezifischer Sicht nicht Teil des Weltwissens sein können, mit dem das Grundlagenmodell ursprünglich trainiert wurde.

Im unten angefügten Bildausschnitt werden bspw. Informationen aus dem amtlichen Lohnsteuerhandbuch 2023 durch ein RAG-System extrahiert. Dabei wurde insbesondere nach der Besteuerung von Künstlern gefragt, also eines sehr speziellen Teilbereiches. Die Antwort enthält dabei sowohl das im Lohnsteuerhandbuch enthaltene Detailwissen als auch Quellenangaben zu den entsprechenden Abschnitten.

Was ist bei der Besteuerung von Künstlern in Deutschland zu beachten?

Besteuerung von Künstlern in Deutschland

Die Besteuerung von Künstlern in Deutschland richtet sich nach verschiedenen Kriterien, insbesondere danach, ob es sich um eine selbständige oder nichtselbständige Tätigkeit handelt und ob der Künstler unbeschränkt oder beschränkt einkommensteuerpflichtig ist. Im Folgenden werden die wesentlichen Aspekte der Besteuerung von Künstlern dargelegt.

Abgrenzung zwischen selbständiger Tätigkeit und nichtselbständiger Arbeit

Für die Annahme einer nichtselbständigen Arbeit sind die in § 1 LStDV aufgestellten Merkmale maßgebend. Es kommt darauf an, ob die tätige Person in der Betätigung ihres geschäftlichen Willens unter der Leitung eines Arbeitgebers steht oder in den geschäftlichen Organismus des Arbeitgebers eingegliedert und dessen Weisungen zu folgen verpflichtet ist. Die formelle vertragliche Gestaltung ist dabei weniger entscheidend als die Durchführung der getroffenen Vereinbarung [Anhang 10 I].

Besteuerung der Einkünfte aus nichtselbständiger Arbeit bei beschränkt einkommensteuerpflichtigen Künstlern

Beschränkt einkommensteuerpflichtige Künstler, die in Deutschland kurzfristig abhängig beschäftigt sind, können unter bestimmten Voraussetzungen einem Pauschsteuersatz unterliegen. Dies betrifft Künstler, die als gastspielverpflichtete Künstler bei Theaterbetrieben, als freie Mitarbeiter für den Hörfunk oder Fernsehfunke oder als Mitarbeiter in der Film- und Fernsehproduktion tätig sind [Anhang 10 II].

Pauschale Erhebung der Lohnsteuer

Bei kurzfristig beschäftigten Künstlern kann die Lohnsteuer pauschal erhoben werden, um die steuerliche Erfassung der Einkünfte zu vereinfachen. Die pauschale Lohnsteuer beträgt 20 % der Einnahmen, wenn der Künstler die Lohnsteuer trägt. Übernimmt der Arbeitgeber die Lohnsteuer und den Solidaritätszuschlag, so beträgt die Lohnsteuer 25,35 % der Einnahmen. Der Solidaritätszuschlag beträgt zusätzlich jeweils 5,5 % der Lohnsteuer [Anhang 10 II].

Lohnsteuerbescheinigung

Eine Lohnsteuerbescheinigung ist vom Arbeitgeber nur auf besonderen Antrag des Künstlers zu erteilen. Sie ist erforderlich, wenn der Künstler Staatsangehöriger eines Mitgliedstaates der Europäischen Union oder von Island, Liechtenstein oder Norwegen ist, in einem dieser Staaten ansässig ist und eine Veranlagung zur Einkommensteuer beantragen will. Die pauschale Lohnsteuer ist auf die veranlagte Einkommensteuer anzurechnen [Anhang 10 II].

Abbildung 15: Antworten aus klassischem RAG-System, wts pIAIground

Inwiefern eine Weiterentwicklung des Systems mit Hilfe von KG eine zusätzliche, substantielle Verbesserung der Antworten liefert, hängt immer auch vom konkreten Anwendungsfall ab. Es ist jedoch davon auszugehen, dass sachbezogene Beziehungen und Hierarchien zwischen Dokumenten kontextbewusst verarbeitet werden können. Dadurch können sie insbesondere für den Steuerbereich eine interessante Weiterentwicklung des RAG-Systems darstellen.

5 Schlussfolgerung

Mit Blick auf die Entwicklungen im Bereich der KI und die Studienergebnisse aus dem Jahr 2017 lassen sich klare Trends ableiten, was die Möglichkeiten und Potenziale für den Einsatz von KI-Modellen bzw. KI-Anwendungen im Steuerbereich betrifft. Sei es, um die bisherigen Arbeitsabläufe innerhalb von Steuerabteilungen effizienter zu gestalten oder komplette Prozesse zu automatisieren. Immer unter Berücksichtigung aktueller Entwicklungen des Steuermarktes, der technologischen Möglichkeiten und der Erfüllung datenschutzrechtlicher Anforderungen.

Hierfür wurden in der vorliegenden Studie von 2024 in enger Zusammenarbeit zwischen WTS und dem DFKI entsprechende Schlüsseltechnologien identifiziert und spezifische Anwendungsszenarien skizziert. Zudem wurden praxisnah die Potenziale für im Steuerbereich relevante Anwendungsfelder beschrieben und die dort vorhandenen Divergenzen und Problemstellungen erörtert. Darüber hinaus liefert die Studie konkrete Beispiele für die prototypische Umsetzung konkreter KI-Methoden, um deren Einsatzpotenziale für die einzelnen Steuerbereiche besser bewerten zu können.

Insbesondere die generativen Sprachmodelle bieten dabei ein enormes Potenzial, um z. B. unternehmensinternes Wissen in Kombination mit steuerlichen Gesetzesgrundlagen zur Beantwortung von steuerlichen Fragestellungen zu nutzen. Dennoch zeigt sich in den einzelnen Experten-Interviews, dass die großen technischen Entwicklungen bisher (noch) nicht breitflächig in der Praxis angekommen sind. Sei es aufgrund fehlenden Know-hows, nicht vorhandener Budgets oder Bedenken, was den Datenschutz betrifft. Und das trotz deutlich günstigerer Rahmenbedingungen und performanter Eigenschaften, die den Einsatz von KI im Steuerbereich zunehmend leichter gestalten.

So viel steht laut Aussagen der interviewten Experten der einzelnen Steuerbereiche fest: KI-Systeme werden eher früher als später Einzug halten und den längst fälligen Wandel weiter vorantreiben. Dabei werden vor allem die aktuellen KI-Technologien der generativen Sprachmodelle zum Zug kommen. Was nicht heißt, dass die länger erforschten Methoden wie bspw. Klassifikation und Anomalieerkennung an Relevanz verlieren. Ganz im Gegenteil, je nachdem welcher Anwendungsfall zu bearbeiten ist, müssen die entsprechenden Lösungen und Anwendungsfälle individuell ermittelt werden.

Vor allem aber setzt der Einsatz von KI-Systemen eine entsprechende Digitalisierung der Steuerfunktion und das Vorhandensein qualitativ hochwertiger Daten voraus. Das schließt die Weiterbildung bzw. das Einstellen von Mitarbeitern mit entsprechenden Skills mit ein, um mit KI-Anwendungen die täglichen Herausforderungen erfolgreich bewältigen zu können. Denn selbst die beste KI-Anwendung kann vorhandene Personalengpässe nur im begrenzten Maße entgegenwirken. In diesem Zusammenhang wird deutlich, dass der Faktor Mensch im Steuerbereich auch in Zukunft immer noch eine wichtige Rolle spielen wird. Sei es als Tax Prompter mit dem Wissen, der KI-Anwendung die richtigen Fragen zu stellen oder als prüfende Instanz, um etwaige Fehler auszugleichen. Im Zuge dessen müssen auch die Beratungsleistungen umgestellt oder zumindest angepasst werden.

Die technischen Voraussetzungen sind also mehr als nur gegeben. Erste KI-Methoden und Prototypen sind mittlerweile am Markt verfügbar, die einen relativ niedrigschwelligen Einstieg ermöglichen. Wichtigste Aufgabe der Steuerabteilungen bzw. Beratungsunternehmen ist es nun, die neuen Voraussetzungen schnellstmöglich in einsatzfähige Produkte zu übersetzen und gezielt anzuwenden. Dabei darf das Kosten-Nutzen-Verhältnis nicht aus dem Fokus geraten. So ist es zu vermeiden, KI-Lösungen zu entwickeln, die am Ende keinerlei Mehrwert bringen und nur als „technologisches Feigenblatt“ dienen. Hinzu kommt eine durchdachte und von Anfang an über alle Bereiche hinweg geplante KI-Strategie, die begleitet durch ein konsequentes Change Management alle Mitarbeiter im Unternehmen abholt. Sei es, um deren Vorbehalte gegenüber neuen Technologien bzw. Prozessen abzubauen oder besser noch deren Affinität für die Nutzung von KI-Anwendungen zu fördern. Nur so gelingt der Übergang in ein neues Zeitalter der Steuerberatung.

Glossar

Accuracy

Accuracy beschreibt eine Metrik, die den Anteil der Fälle (sowohl richtig Positive als auch richtig Negative) im Verhältnis zur Gesamtzahl aller Fälle misst. Die Metrik gibt Rückschlüsse auf die Korrektheit des Gesamtmodells, ist insbesondere bei unausgewogene Klassenverteilungen nicht immer repräsentativ.

Anomalieerkennung

Bei der Erkennung von Anomalien geht es um die Identifizierung von Datenpunkten, Ereignissen oder Beobachtungen, die erheblich von dem erwarteten Muster in einem Datensatz abweichen. Diese Ausreißer deuten oft auf Unregelmäßigkeiten, Betrug oder Systemstörungen hin. Mithilfe verschiedener Techniken der Anomalieerkennung sollen verborgene Muster aufgedeckt und mögliche schädliche Vorfälle vorhergesagt werden.

Bestärkendes Lernen

Bezeichnet eine Variante des maschinellen Lernens, bei dem ein Agent lernt, seine Leistung durch Interaktion mit seiner Umgebung zu verbessern. Der Agent lernt anhand von Feedback, das durch eine Belohnungsfunktion beschrieben wird und beschreibt, wie gut oder schlecht die Aktion des Agenten war. Durch den Prozess des Ausprobierens und Lernens aus Erfahrungen entwickelt der Agent somit Strategien, um komplexe Aufgaben zu lösen oder spezifische Ziele zu erreichen.

Bias

Der Begriff *Bias* wird allgemein als eine Form der Verzerrung verstanden. In der Psychologie bezieht es sich auf Einstellungen oder Stereotype, die unsere Wahrnehmung der Umwelt, sowie unsere Entscheidungen und Aktionen positiv oder negativ beeinflussen können. Im Kontext von ML bezeichnet Bias einen Fehler, der in den Trainingsdaten vorhanden ist. Diese Verzerrungen können sich auf verschiedene Weisen manifestieren (z. B. durch diskriminierende Sprache, Falschinformationen, etc.) und in jeder Phase des ML-Lebenszyklus entstehen (z. B. bei der Datenerhebung, Training und Evaluation).

Big Data

Big Data bezeichnet große und komplexe Datensätze, die von herkömmlichen Datenverarbeitungssystemen nicht effizient verarbeitet werden können. Sie zeichnen sich durch ihr Volumen, ihre Vielfalt, ihre Geschwindigkeit und Wahrhaftigkeit aus und umfassen strukturierte, halbstrukturierte und unstrukturierte Daten. Die Analyse von Big Data liefert Erkenntnisse, Muster und Trends, die die Entscheidungsfindung unterstützen und Innovation vorantreiben können.

Concept Drift

Concept Drift beschreibt die Veränderung der Zuordnung von Eingabevariablen (Merkmale) und der Zielvariablen (Label). Hierbei kommt es zu Änderungen in der Beziehung, d.h. der Art und Weise, wie sich diese Merkmale darauf beziehen, eine bestimmte Klasse zu beschreiben. Dadurch kann es zu einer Verschlechterung der Leistung und Genauigkeit eines ML-Modells kommen.

Data Drift

Data Drift beschreibt das Problem, wenn es zu einer Veränderung der Verteilung der Eingabevariablen (Merkmale) kommt, jedoch die bedingte Wahrscheinlichkeitsverteilung der Zielvariablen (Label) gleich bleibt (Covariate Shift). Ändern sich die Eingabevariablen (z. B. neue Ausprägungen, längere Texte usw.), kann es sein, dass die Verknüpfung zur ursprünglichen Klasse verschwindet. Ähnliches gilt auch für den Fall, dass sich die Verteilung der Zielvariablen ändert, jedoch für die Eingabevariablen nicht (Label Shift). Entsprechend kann es zu einer Verschlechterung der Leistung und Genauigkeit eines ML-Modells kommen.

Data Leakage

Data Leakage beschreibt ein Konzept, in der ein ML-Modell auf Daten trainiert wird, auf die es eigentlich in der realen Anwendung keinen Zugriff hätte. Das führt dazu, dass das ML-Modell im Rahmen der Evaluation schummelt, indem es etwas vorhersagt, das bereits im Training schon bekannt war, und somit sehr gut abschneidet. In der praktischen Anwendung jedoch wird es auf neuen, ungesehenen Daten tendenziell schlechter abschneiden.

Deep Learning

Deep Learning ist ein Teilbereich des maschinellen Lernens, bei dem neuronale Netze mit mehreren Schichten verwendet werden. Durch das Training mit großen Datenmengen sind Deep-Learning-Modelle in der Lage, automatisch komplexe Merkmale und Darstellungen aus Rohdaten zu erlernen und zu extrahieren, wodurch eine bemerkenswerte Genauigkeit insbesondere bei der Bearbeitung komplexer Aufgaben wie der Bild- und Spracherkennung und Verarbeitung natürlicher Sprache erreicht werden kann.

Entscheidungsbäume

Entscheidungsbäume (eng. Decision Trees) beschreiben einen nichtparametrischer überwachter Lernalgorithmus, der durch eine hierarchische Aufteilung der Datenpunkte versucht, Entscheidungsregeln zu erlernen, um Daten zu klassifizieren oder kontinuierliche Werte vorherzusagen.

F1-Score

Beschreibt ein Maß für die Genauigkeit eines Modells, das sowohl die *Precision* als auch den *Recall* berücksichtigt. Der F1-Score ist definiert als der harmonische Mittelwert, der sowohl die Identifizierung von positiven Fällen als auch die Abdeckungsgrad aller tatsächlichen positiven Fällen ausgleicht. Die Metrik findet häufig in Anwendungsfällen mit unausgewogenen Datensätzen Anwendung.

Fairness

Fairness bezieht sich auf die durch das Training erworbenen Eigenschaften eines KI-Modells und bezeichnet im Kontext der Entscheidungsfindung das Vorhandensein von Vorurteilen oder Bevorzungen. Ähnlich wie beim Bias ist das Verhalten des ML-Modells auf menschliche Vorurteile und Stereotypisierung durch sensible Datenattribute zurückzuführen. Diese können absichtlich oder unabsichtlich auftreten.

Falsch-Positiv-Rate

Die Falsch-Positiv-Rate (engl. false positive rate) gibt den Anteil der fälschlicherweise als positiv klassifizierten Objekte an der Gesamtheit der tatsächlich negativen Objekte an. Im prüferischen Kontext beschreibt die Metrik somit die Effizienz des zugrundeliegenden Modells.

Halluzination

Large Language Models können dazu neigen, sogenannte „Halluzinationen“ (Falschinformationen) zu erzeugen. Konkret bedeutet das, dass Informationen generiert werden, die überzeugend klingen, aber falsch oder irreführend sind. Dieses Problem entsteht dadurch, dass das Modell Muster in den Trainingsdaten „sieht“, die nicht unbedingt der Realität entsprechen. Entsprechend sind Halluzinationen besonders problematisch für Anwendungsbereiche, in denen der Nutzer auf die Genauigkeit der bereitgestellten Informationen angewiesen ist.

Hyperparameter

Hyperparameter bezeichnen anpassbare Parameter, die Einfluss auf das Verhalten eines maschinellen Lernmodells haben. Bei der Optimierung von Hyperparametern geht es darum, die optimalen Parametereinstellungen zu finden, um die Vorhersagegenauigkeit des Modells zu verbessern und gleichzeitig eine Überanpassung zu verhindern.

Internet der Dinge

Unter dem Internet der Dinge versteht man Technologien, mit denen physische und virtuelle Objekte über das Internet miteinander vernetzt werden. Entsprechende Objekte sind i.d.R. mit Sensoren zur Wahrnehmung der Umgebung sowie mit Aktuatoren zur Ausführung bestimmter Aktionen ausgestattet. Durch diese Vernetzung findet ein stetiger Informationsaustausch statt, wodurch in meist automatisierter Weise verschiedenste Funktionen durch die Objekte erfüllt werden können.

Konfusionsmatrix

Die Konfusionsmatrix (engl. confusion matrix) dient beim ML zur Bewertung der Leistung eines Klassifizierungsmodells. Sie umfasst sämtliche Ergebnisse der Vorhersage eines Modells für einen Datensatz zusammen und vergleicht die Vorhersagen mit den wahren Bezeichnungen in diesen. Die Konfusionsmatrix umfasst insgesamt vier Kennzahlen: True-Positive (TP), False-Positive (FP), True-Negative (TN) und False-Negative (FN). TP gibt die Anzahl der korrekten positiven Klassifizierungen an. Die TN hingegen die Anzahl der korrekten negativen Klassifizierungen (Ablehnungen). FP beschreibt die Anzahl der falschen positiven Klassifizierungen (Fehlalarm) wohingegen FN die positiven Klassifizierungen, die fälschlicherweise als negativ klassifiziert wurden, angibt.

Große Sprachmodelle

Sehr große und umfangreiche neuronale Netze, die in der Regel aus Milliarden von Parametern bestehen und auf einer sehr großen Menge an Datensätzen trainiert werden, um menschenähnlichen Text zu verstehen, zu interpretieren und zu erzeugen. Große Sprachmodelle, insbesondere solche, die auf der Transformer-Architektur basieren, haben mehrere Vorteile gegenüber anderen Netzwerkarchitekturen wie Recurrent Neural Networks (RNN) und Long Short-Term Memory (LSTM) Netzwerke, da Texteingaben parallel verarbeitet und Kontext mitberücksichtigt wird.

LIME

LIME ist ein lokales Verfahren zur Erklärung von Vorhersagen für eine Eingabe. Es benötigt darüber hinaus kein „Wissen“ über Details des zu erklärenden Modells (bspw. dessen Architektur), wodurch dieser Ansatz für unterschiedlichste Modellarten ohne tiefgehende Anpassungen herangezogen werden kann. Die zentrale Vorgehensweise bei diesem Ansatz besteht darin, dass die ursprünglichen Trainingsdaten manipuliert und dadurch bedingte „Verhaltensänderungen“ des zu erklärenden Modells gemessen werden. Konkret wird basierend auf den manipulierten Daten und den neu erhaltenen Vorhersagen ein weiteres erklärbares Modell (z. B. ein lineares Modell) trainiert, um die resultierenden Abweichungen zu erfassen.

Maschinelles Lernen

Maschinelles Lernen (ML) ist ein Teilbereich der KI, das sich mit der Entwicklung mathematisch-statistischer Verfahren beschäftigt, um Algorithmen die Fähigkeit zu verleihen, anhand von Daten selbstständig zu "lernen", ohne dass sie explizit programmiert werden müssen. Die Art und Weise, wie Algorithmen lernen, kann in drei Klassen unterschieden werden: Überwachtes Lernen, Unüberwachtes Lernen und Bestärkendes Lernen.

Model Drift

Bezeichnet die Veränderung der Leistung und Genauigkeit eines ML-Modells im Laufe der Zeit aufgrund Veränderungen in den Eingabedaten. Allgemein wird zwischen zwei Arten von Veränderung unterschieden, nämlich *Concept Drift* und *Data Drift*.

Natural Language Processing

Teilgebiet der KI, das sich mit der Verarbeitung und dem Verstehen natürlicher Sprache beschäftigt. Dazu kombiniert es Konzepte aus den Bereichen der Computerlinguistik, Kognitionswissenschaft und Informatik., Dadurch lässt sich Sprache (geschrieben, gesprochen, visuell) interpretieren, um Aufgaben wie Übersetzungen, Sentiment Analysen, Textgenerierung und Chatbots zu unterstützen.

Optical Character Recognition (OCR)

Bezeichnet eine Klasse von verschiedenen Technologien zur Übersetzung einer optischen oder auch bildhaften Darstellung eines Schriftzeichens in seine maschinenlesbare Form. Dies kann neben den Symbolen in eingescannten maschinellen Texten auch handschriftlich ausgefüllte Formulare oder ganze schreiben umfassen. OCR umfasst allerdings lediglich den Schritt der direkten Umwandlung einzelner Symbole und Symbolketten. Eine Zuordnung der Symbolketten zu einzelnen Kategorien, wie Formularfeldern oder spezifischen Nummern wird durch OCR nicht abgedeckt, sondern nur in Kombination mit der Technologie Informationsextraktion.

Parameter-Efficient Fine-Tuning (PEFT)

Beschreibt Techniken, vortrainierte Modelle mit einer minimalen Anzahl von zusätzlichen oder angepassten Parametern für eine spezifische Aufgabe anzupassen. Ziel ist es, die Leistungsfähigkeit von großen Sprachmodellen zu nutzen, ohne alle Modellparameter anzupassen. Dadurch sind PEFT-Techniken deutlich ressourceneffizienter und schneller umzusetzen, ohne signifikant an Modellperformanz zu verlieren.

Precision

Die Präzision (engl. precision) beschreibt, wie viele richtig vorhergesagte positive Fälle es unter allen vorhergesagten positiven Fällen (richtig und falsch) gibt. Die Metrik gibt einen spezifischen Einblick, wie gut das Modell Transaktionen der positiven Klasse klassifiziert und falsch-positive Fälle vermieden werden.

Retrieval-Augmented Generation (RAG)

Ein RAG-System kombiniert ein großes Sprachmodell mit einem Retrieval-System, um relevante Informationen aus einer großen Datenbank oder einem Dokumentenkorpus als Kontext für die Lösung der zugrundeliegenden Aufgabe nutzen. Diese Methode ermöglicht es, präzise und informierte Antworten zu generieren, indem sie spezifische Informationen zusätzlich bereitstellt, die über das ursprüngliche Trainingsdatenset des großen Sprachmodells hinausgehen.

Recall

Der Recall misst wie viele richtig vorhergesagte positive Fälle es unter allen tatsächlichen positiven Fällen gibt und beschreibt somit den Abdeckungsgrad der Vorhersage, wie viele der tatsächlichen positiven Fälle identifiziert werden (Vermeidung von falsch positiven Fällen).

Richtig-Positiv-Rate

Die Richtig-Positiv-Rate (engl. true positive rate) gibt den Anteil der richtig positiv klassifizierten Objekte an der Gesamtheit der tatsächlich positiven Objekte an. Sie gibt an, wie sensitiv das Modell auf positive Objekte reagiert.

RLHF

Reinforcement Learning from Human Feedback (RLHF) bezeichnet einen Ansatz im maschinellen Lernen, bei dem ein KI-Modell wie bspw. GPT-4 durch menschliches Feedback statt ausschließlich durch vordefinierte Belohnungsfunktionen trainiert werden. Dies ermöglicht es dem Modell, komplexe Aufgabenstellungen zu erfassen und menschliches Verhalten oder Präferenzen besser zu verstehen.

SHAP

SHAP ist ein lokales XAI-Verfahren und basiert auf grundlegenden Konzepten der Spieltheorie. Das Ziel von SHAP ist, den Einfluss einzelner Datenattribute auf die Vorhersagen zu identifizieren. Dies gelingt durch Betrachtung von permutierten Teilmengen von Datenattributen und den dadurch bedingten Auswirkungen auf die Vorhersagen. Konkret werden Teilmengen abgeleitet, in denen ursprüngliche Werte gewisser Attribute durch randomisierte Werte ersetzt werden. Die Idee dahinter ist, dass solche randomisierten Ausprägungen keine Vorhersagekraft besitzen. Bei SHAP wird für jedes Attribut der sog. Shapley Value berechnet, welcher u. a. auf der Differenz zwischen den Vorhersagen mit und ohne Betrachtung des zugrundeliegenden Attributes basiert. Durch diese Differenzen können schlussendlich Einflüsse der einzelnen Attribute erhoben werden.

Transfer Learning

Bezeichnet einen ML-Ansatz, bei dem ML-Modelle die auf eine spezifische Aufgabe trainiert wurden, auf eine artverwandte Aufgabe angewandt werden. Transfer Learning ermöglicht es das Wissen eines Modells zu nutzen, um eine andere Aufgabe besser zu bewältigen. Dies betrifft insbesondere Aufgaben, die ein gewisses "Allgemeinverständnis" benötigen, wie bspw. Die Verarbeitung natürlicher Sprache oder Computer Vision.

Unüberwachtes Lernen

Bezeichnet einen ML-Ansatz, bei dem Algorithmen Muster und Strukturen in Daten erkennen, ohne dass explizite Anweisungen oder gelabelte Daten vorgegeben werden. Diese Technik wird häufig eingesetzt, um verborgene Zusammenhänge in großen und ungelabelten Datensätzen zu entdecken.

Überwachtes Lernen

Bezeichnet einen ML-Ansatz im Machine Learning, bei dem ein Modell anhand von gelabelten Trainingsdaten trainiert werden, um eine Beziehung zwischen Eingabedaten und ihren zugehörigen Ausgaben (Labels) zu erlernen. Das Ziel besteht darin, auf Grundlage der Trainingsdaten Vorhersagen oder Entscheidungen für neue, ungesehene Daten zu treffen.

Literatur

- [1] P. Fettke, G. Herzog, J. Lahann, H. Maus, G. Neumann, und T. Niesen, „Künstliche Intelligenz im Steuerbereich. Innovationsstudie zur Digitalisierung und den Potentialen Künstlicher Intelligenz im Bereich Steuer“, München, 2017.
- [2] A. Vaswani u. a., „Attention is All You Need“, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, S. 6000–6010.
- [3] J. Devlin, M.-W. Chang, K. Lee, und K. Toutanova, „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“, in *Proceedings of the 2019 Conference of the North*, 2019, S. 4171–4186.
- [4] A. Radford und K. Narasimhan, „Improving Language Understanding by Generative Pre-Training“, 2018.
- [5] O. Sharir, B. Peleg, und Y. Shoham, „The Cost of Training NLP Models: A Concise Overview“, *arXiv:2004.08900*. 2020.
- [6] N. Maslej u. a., „The AI Index 2023 Annual Report“, Stanford University, Stanford, CA, 2023.
- [7] Taxy.io, „Besteht ChatGPT steuerfachliche Prüfungen? Modell: GPT-4 - Update zur Studie mit GPT 3.5“. Aachen, 2023.
- [8] G. Kümmerle, „Steuer-KI von Taxy.io und Otto-Schmidt-Verlag mit neuer Technologie“, *Juve Steuermarkt*, 2023. [Online]. Verfügbar unter: <https://www.juve-steuermarkt.de/branche/steuer-ki-von-taxy-io-und-otto-schmidt-verlag-mit-neuer-technologie/>.
- [9] J. McCarthy, M. Minsky, N. Rochester, und C. E. Shannon, „A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE“, *Website*, 1996. [Online]. Verfügbar unter: <https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
- [10] W. S. McCulloch und W. Pitts, „A logical calculus of the ideas immanent in nervous activity“, *Bull. Math. Biophys.*, Bd. 5, Nr. 4, S. 115–133, 1943.
- [11] F. Rosenblatt, „The Perceptron. A Perceiving And Recognizing Automation“, Buffalo, N.Y., 1957.
- [12] I. J. Goodfellow u. a., „Generative adversarial nets“, in *Advances in Neural Information Processing Systems*, 2014, Bd. 3, Nr. January, S. 2672–2680.
- [13] R. Eckel, „Microsoft Researchers' Algorithm Sets ImageNet Challenge Milestone“, *Microsoft Research Post*, 2015. [Online]. Verfügbar unter: <https://www.microsoft.com/en-us/research/blog/microsoft-researchers-algorithm-sets-imagenet-challenge-milestone/>.
- [14] C. Czarnecki und P. Fettke, „Robotic process automation“, in *Robotic Process Automation*, C. Czarnecki und P. Fettke, Hrsg. Berlin, Boston: De Gruyter, 2021, S. 3–24.
- [15] W. M. P. van der Aalst, M. Bichler, und A. Heinzl, „Robotic Process Automation“, *Bus. Inf. Syst. Eng.*, Bd. 60, Nr. 4, S. 269–272, Aug. 2018.
- [16] A. Asatiani und E. Penttinen, „Turning robotic process automation into commercial success – Case OpusCapita“, *J. Inf. Technol. Teach. Cases*, Bd. 6, Nr. 2, S. 67–74, 2016.
- [17] X. Zhang und Z. Wen, „Thoughts on the development of artificial intelligence combined with RPA“, *J. Phys. Conf. Ser.*, Bd. 1883, Nr. 1, 2021.
- [18] S. Aguirre und A. Rodriguez, „Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study“, in *Applied Computer Sciences in Engineering*, 2017, S. 65–71.
- [19] P. Fettke und S. Strohmeier, „Chapter 10: HR robotic process automation“, in *Handbook of Research on Artificial Intelligence in Human Resource Management*, Cheltenham, UK: Edward Elgar Publishing, 2022.
- [20] L.-V. Herm, C. Janiesch, H. A. Reijers, und F. Seubert, „From Symbolic RPA to Intelligent RPA: Challenges for

Developing and Operating Intelligent Software Robots", in *Business Process Management*, 2021, S. 289–305.

- [21] J. Viehhauser, „Is Robotic Process Automation Becoming Intelligent? Early Evidence of Influences of Artificial Intelligence on Robotic Process Automation", in *Business Process Management: Blockchain and Robotic Process Automation Forum*, 2020, S. 101–115.
- [22] A. S. Villar und N. Khan, „Robotic process automation in banking industry: a case study on Deutsche Bank", *J. Bank. Financ. Technol.*, Bd. 5, Nr. 1, S. 71–86, 2021.
- [23] B. Kuzniacki, M. Almada, K. Tyliński, und Ł. Górski, „Requirements for Tax XAI Under Constitutional Principles and Human Rights", in *Explainable and Transparent AI and Multi-Agent Systems*, 2022, S. 221–238.
- [24] N. Gnos, M. Schultz, und M. Tropmann-Frick, „XAI in the audit domain : explaining an autoencoder model for anomaly detection", in *Internationale Tagung Wirtschaftsinformatik 2022*, 2022.
- [25] G. Press, „Andrew Ng Launches A Campaign For Data-Centric AI", *Forbes*, 2021. [Online]. Verfügbar unter: <https://www.forbes.com/sites/gilpress/2021/06/16/andrew-ng-launches-a-campaign-for-data-centric-ai/?sh=664bf56374f5>.
- [26] A. Athalye, C. Northcutt, und J. Mueller, „Data-Centric AI vs. Model-Centric AI", *MIT*, 2023. [Online]. Verfügbar unter: <https://dcai.csail.mit.edu/lectures/data-centric-model-centric/>.
- [27] E. Strickland, „Anrew Ng: Unbiggen AI", *IEEE Spectrum*, 2022. [Online]. Verfügbar unter: <https://spectrum.ieee.org/andrew-ng-data-centric-ai>.
- [28] WAYMO, „Cities, freeways, airports: How we've built a scalable autonomous driver", 2022. [Online]. Verfügbar unter: <https://waymo.com/blog/2022/05/howwevebuiltascalableautonomousdriver.html>.
- [29] A. Karpathy, „Building the Software 2.0 Stack", *vimeo*. 2018.
- [30] A. Zewe, „In machine learning, synthetic data can offer real performance improvements", *MIT News*, 2022. [Online]. Verfügbar unter: <https://news.mit.edu/2022/synthetic-data-ai-improvements-1103>.
- [31] I. Shumailov, Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, und R. Anderson, „The Curse of Recursion: Training on Generated Data Makes Models Forget", *arXiv:2305.17493*. 2023.
- [32] N. Al-Sibai, „AI DEVELOPERS ARE ALREADY QUIETLY TRAINING AI MODELS USING AI-GENERATED DATA", 2020. [Online]. Verfügbar unter: <https://futurism.com/the-byte/ai-synthetic-data>.
- [33] J. Jordon u. a., „Synthetic Data -- what, why and how?", *arXiv:2205.03257*. 2022.
- [34] OECD, „Tax Administration 2022: Comparative Information on OECD and other Advanced and Emerging Economies", 2022.
- [35] P. Breuer, D. Hagemeyer, und H. Hürtgen, „Artificial intelligence: How advanced analytics and smart machines will change the way we work", 2018.
- [36] B. Zergibel, „Künstliche Intelligenz und Blockchain-Technologie - Eine Symbiose mit Zukunftspotenzial", *REthinking:Tax*, Bd. 4, S. 9–17, 2022.
- [37] C. Schmidt, „Quo vadis, Finanzverwaltung? Potenziale und Herausforderungen eines künftigen behördlichen KI-Einsatzes", *REthinking:Tax*, Bd. 1, S. 70–81, 2022.
- [38] Bundeszentralamt für Steuern, „Das BZSt nimmt den FAQ-Chatbot ViOIA in Betrieb", 2023. [Online]. Verfügbar unter: https://www.bzst.de/SharedDocs/Kurzmeldungen/DE/2021_Kurmeldungen/2021090_livegang_viola.html.
- [39] Informationstechnikzentrum Bund, „ITZBund realisiert und betreibt Chatbot TinA", 2021. [Online]. Verfügbar unter: https://www.itzbund.de/SharedDocs/Kurzmeldungen/DE/2021-10-22_Chatbot_TinA.html.

- [40] S. Raschka, *Machine Learning – Q and AI*. Lean Publishing, 2023.
- [41] B. Plüster, „LEOLM: EIN IMPULS FÜR DEUTSCHSPRACHIGE LLM-FORSCHUNG“, *LAION*, 2023. [Online]. Verfügbar unter: <https://laion.ai/blog-de/leo-lm/>.
- [42] V. Lialin, V. Deshpande, und A. Rumshisky, „Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning“, *arXiv:2303.15647*. 2023.
- [43] S. Bergmann, „Neue Verrechnungspreisdokumentationspflichten für multinationale Unternehmensgruppen“, *Zeitschrift für Gesellschaftsr.*, Bd. 2016, Nr. 4, S. 147–148, 2016.
- [44] OECD, „Practical Toolkit to Support the Successful Implementation By Developing Countries Of Effective Transfer Pricing Documentation Requirements“, 2021.
- [45] M. Skalický, Š. Šimsa, M. Uříčář, und M. Šulc, „Business Document Information Extraction: Towards Practical Benchmarks“, in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, A. Barrón-Cedeño, G. Da San Martino, M. Dgli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, und N. Ferro, Hrsg. Cham: Springer International Publishing, 2022, S. 105–117.
- [46] C. Houy, M. Hamberg, und P. Fettke, „Robotic Process Automation in Public Administrations“, in *Digitalisierung von Staat und Verwaltung*, 2019, S. 62–74.
- [47] L. Chiticariu, Y. Li, und F. R. Reiss, „Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!“, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, S. 827–832.
- [48] B. Oral, E. Emekligil, S. Arslan, und G. Eryigit, „Information Extraction from Text Intensive and Visually Rich Banking Documents“, *Inf. Process. Manag.*, Bd. 57, Nr. 6, Nov. 2020.
- [49] J. Hanken und G. Kleinhietaß, *Verrechnungspreise. Im Spannungsfeld von Controlling und Steuern*, 1. Aufl. Freiburg: Haufe-Lexware GmbH & Co. KG, 2014.
- [50] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, und B. Xu, „Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling“, in *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, 2016, S. 3485–3495.
- [51] M. V. Koroteev, „BERT: A Review of Applications in Natural Language Processing and Understanding“, *arXiv:2103.11943*. 2021.
- [52] M. Schreyer, T. Sattarov, D. Borth, A. Dengel, und B. Reimer, „Detection of Anomalies in Large Scale Accounting Data using Deep Autoencoder Networks“, *arXiv:1709.05254*. 2017.
- [53] M. Braun, J. Köppe-Karkutsch, und T. Hoppe, „Generative AI ist gekommen um zu bleiben – Meet your new Coworkers!“, *beck.digitax*, Bd. 46, 2024.
- [54] N. Kandpal, H. Deng, A. Roberts, E. Wallace, und C. Raffel, „Large Language Models Struggle to Learn Long-Tail Knowledge“, in *Proceedings of the 40th International Conference on Machine Learning*, 2023, Bd. 202, S. 15696–15707.
- [55] P. Lewis u. a., „Retrieval-augmented generation for knowledge-intensive NLP tasks“, in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [56] Y. Gao u. a., „Retrieval-Augmented Generation for Large Language Models: A Survey“, *arXiv:2312.10997*. 2023.
- [57] M. Kang, J. M. Kwak, J. Baek, und S. J. Hwang, „Knowledge Graph-Augmented Language Models for Knowledge-Grounded Dialogue Generation“, *arXiv:2206.03971*, 2023.
- [58] A. Erdl, „Unifying LLMs & Knowledge Graphs for GenAI: Use Cases & Best Practices“, *Neo4j*, 2023. [Online]. Verfügbar unter: <https://neo4j.com/blog/unifying-llm-knowledge-graph>.

- [59] L. Bellomarini, E. Sallinger, und S. Vahdati, „Chapter 2 Knowledge Graphs : The Layered“, in *Knowledge Graphs and Big Data Processing*, Bd. 12072 LNCS, Cham: Springer, 2020, S. 20–34.
- [60] M. Kejriwal, *Domain-specific knowledge graph construction*. Cham: Springer, 2019.
- [61] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, und X. Wu, „Unifying Large Language Models and Knowledge Graphs: A Roadmap“, *IEEE Trans. Knowl. Data Eng.*, 2024.