

MARCO SCHREYER

ANITA GIERBL

T. FLEMMING RUUD

DAMIAN BORTH

# STICHPROBENAUSWAHL DURCH DIE ANWENDUNG VON KÜNSTLICHER INTELLIGENZ

## Lernen repräsentativer Stichproben aus Journalbuchungen in der Prüfungspraxis\*

**Verfahren der künstlichen Intelligenz (KI) etablieren sich zunehmend als wertvolles Werkzeug in der Prüfung. Der nachfolgende Beitrag soll die praktische Einsatzmöglichkeit des Deep Learnings, eines derzeit erfolgreichen Teilgebiets der KI, für das Lernen repräsentativer Stichproben aus umfangreichen Journalbuchungen aufzeigen.**

### 1. STICHPROBENBASIERTE PRÜFUNGSHANDLUNGEN

Das Vertrauen in die finanzielle Berichterstattung von Unternehmen stellt ein grundlegendes Fundament [1] der Entscheidungsfindung unterschiedlichster Adressaten, wie z. B. Investoren, Verwaltungsrat und Lieferanten, dar. In diesem Kontext sind die Prüfungshandlungen von Wirtschaftsprüferinnen und Wirtschaftsprüfern darauf ausgerichtet, eine hinreichende Sicherheit zu gewährleisten, dass ein Jahresabschluss keine wesentlichen Falschdarstellungen aufweist (PS 701). Zugleich ist es aus wirtschaftlichen Gründen selten zweckmässig, die finanzielle Berichterstattung einer lückenlosen Prüfung zu unterziehen [2]. Eine Zielsetzung des prüferischen Vorgehens besteht deshalb darin, möglichst effizient ausreichende Prüfungssicherheit über ein Prüffeld durch geeignete Prüfnachweise zu erlangen. Dies kann im Rahmen der Prüfung bspw. durch eine stichprobenbasierte Prüfung von «buchungsintensiven» Bilanz- und Erfolgsrechnungspositionen eines Unternehmens erfolgen. Heutzutage erfassen eine Vielzahl von Unternehmen umfangreiche Mengen Journalbuchungen (engl. Journal Entries) digital in ERP-Systemen [3]. *Abbildung 1* zeigt schemenhaft die Erfassung solcher Journalbuchungen innerhalb eines ERP-Systems.

Zugleich finden, bedingt durch die rasanten Fortschritte der KI, zunehmend KI-unterstützte Verfahren Eingang in die Prüfung [4, 5, 6, 7, 8]. Diese Entwicklungen werfen die Frage auf, ob sich solche Verfahren auch für Zwecke der re-

präsentativen Stichprobenauswahl eignen. Der folgende Artikel stellt die Anwendung eines neuartigen Verfahrens der KI, sog. Vector Quantised-Variational Autoencoder (VQ-VAE), für die Stichprobenauswahl vor. Anhand ausgewählter Ergebnisse einer empirischen Studie wird dargestellt, wie das Erlernen repräsentativer Stichproben aus umfangreichen Journalbuchungen automatisiert gelingt. Solche Verfahren können für die Einschätzung der Merkmale einer zu prüfenden Grundgesamtheit einen Mehrwert für Prüferinnen und Prüfer stiften.

### 2. STICHPROBENAUSWAHLVERFAHREN UND STICHPROBENRISIKO

Im Rahmen der Durchführung von Einzelfall- und Funktionsprüfungen können Prüferinnen und Prüfer Stichprobenverfahren einsetzen (vgl. PS 530). Formal ist eine Stichprobenprüfung definiert, als die Auswahl und Evaluation von weniger als 100 Prozent der prüfungsrelevanten Grundgesamtheit (PS 530.5a). Zugleich birgt die Stichprobenprüfung jedoch ein Stichprobenrisiko, welches die Wahrscheinlichkeit approximiert,

*«dass die stichprobenbasierte Schlussfolgerung des Abschlussprüfers von der Schlussfolgerung abweicht, wenn die vollständige Grundgesamtheit Gegenstand derselben Prüfungshandlung wäre» (PS 530.5c).*

Um das Stichprobenrisiko zu minimieren, sind Prüferinnen und Prüfer aufgefordert, eine repräsentative Stichprobenaus-

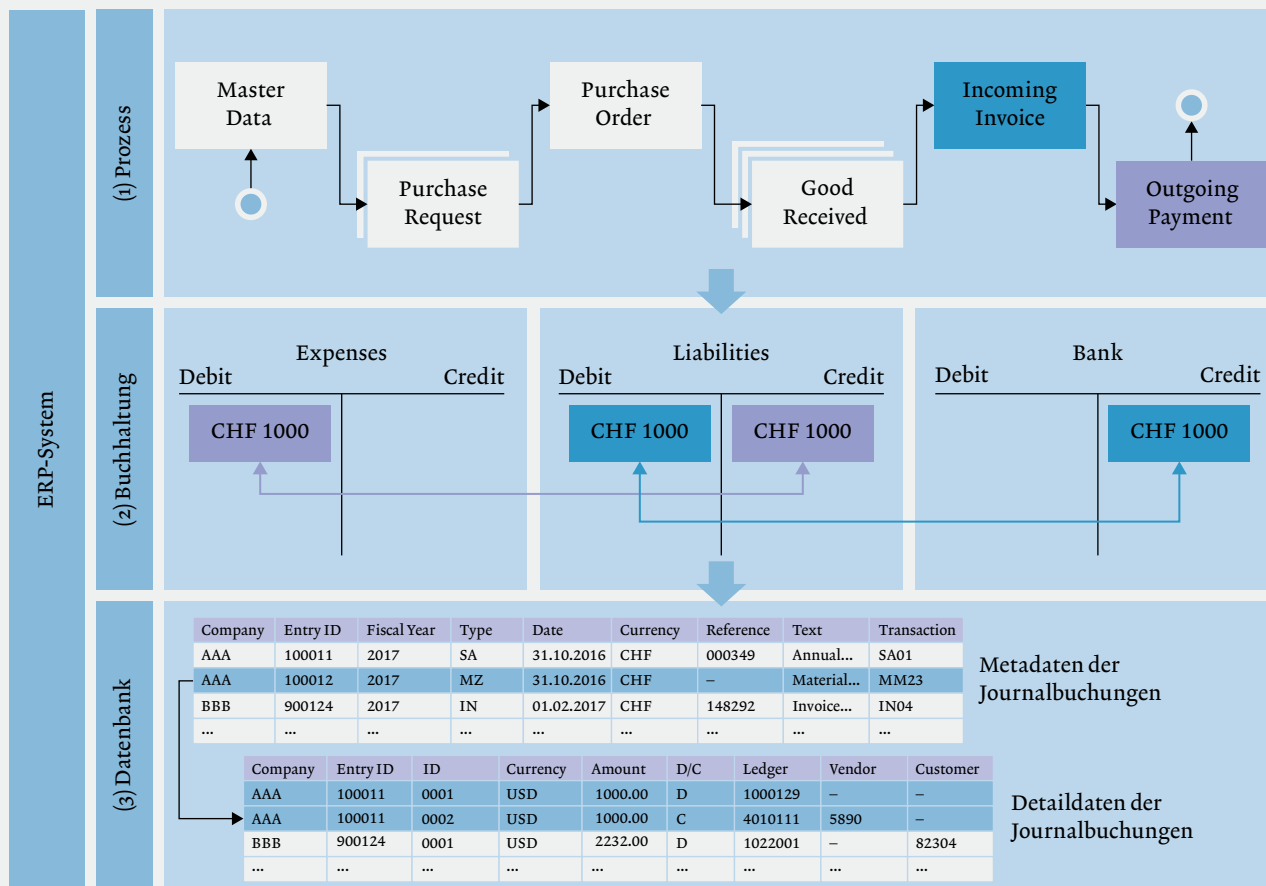


MARCO SCHREYER,  
WISSENSCHAFTLICHER  
MITARBEITER, LEHR-  
STUHL FÜR KÜNSTLICHE  
INTELLIGENZ UND  
MASCHINELLES LERNEN,  
INSTITUTE OF COMPUTER  
SCIENCE (ICS),  
UNIVERSITÄT ST. GALLEN



ANITA GIERBL,  
DR. OEC. HSG,  
WISSENSCHAFTLICHE  
MITARBEITERIN,  
UNIVERSITÄT ST. GALLEN,  
FACHASSISTENZ  
SWISS GAAP FER,  
AUDIT, PWC SCHWEIZ

Abbildung 1: SCHEMATISCHE ERFASSUNG VON JOURNALBUCHUNGEN AUF VERSCHIEDENEN SEMANTISCHEN ABSTRAKTIONSNIVEAUS INNERHALB EINES ERP-SYSTEMS



wahl zu treffen (PS 530.7). Repräsentativ bedeutet, die Stichprobe so auszuwählen, dass diese die strukturellen Merkmale der Grundgesamtheit widerspiegelt und keine systematischen Verzerrungen aufweist [9].

Anhand des PS 530 werden für die Stichprobenprüfung

- die zufallsgesteuerte Auswahl,
- die systematische Auswahl und
- die Auswahl aufs Geratewohl unterschieden (vgl. Abbildung 2). Die zufallsgesteuerte Auswahl gliedert sich weiter in statistische und nichtstatistische Stichprobenverfahren. Im Kontext der statistischen Verfahren erfolgt die Stichprobenauswahl anhand objektiverer Wahrscheinlichkeiten [11]. Im Kontext der nichtstatistischen Verfahren erfolgt die Auswahl subjektiv, d. h. auf der Grundlage des professionellen

Ermessens der Prüferin oder des Prüfers [12]. Die systematische Auswahl ist gekennzeichnet durch vordefinierte Stichprobenintervalle. Im Gegensatz hierzu erfolgt die Auswahl aufs Geratewohl ohne strukturiertes Verfahren.

Prüferinnen und Prüfern obliegt es im Rahmen des Prüfungsvorgehens, selbst festzulegen durch welches Verfahren eine hinreichende Prüfungssicherheit über die zu prüfende Grundgesamtheit eines Prüffelds erlangt werden kann [13]. Für die Auswahl eines geeigneten Auswahlverfahrens sind Prüferinnen und Prüfer jedoch angehalten, die Merkmale der Grundgesamtheit vorab einzuschätzen, z. B. im Hinblick auf Homogenität und strukturelle Ausprägungen. Auf der Grundlage einer solchen Einschätzung beurteilt die Prüferin bzw.

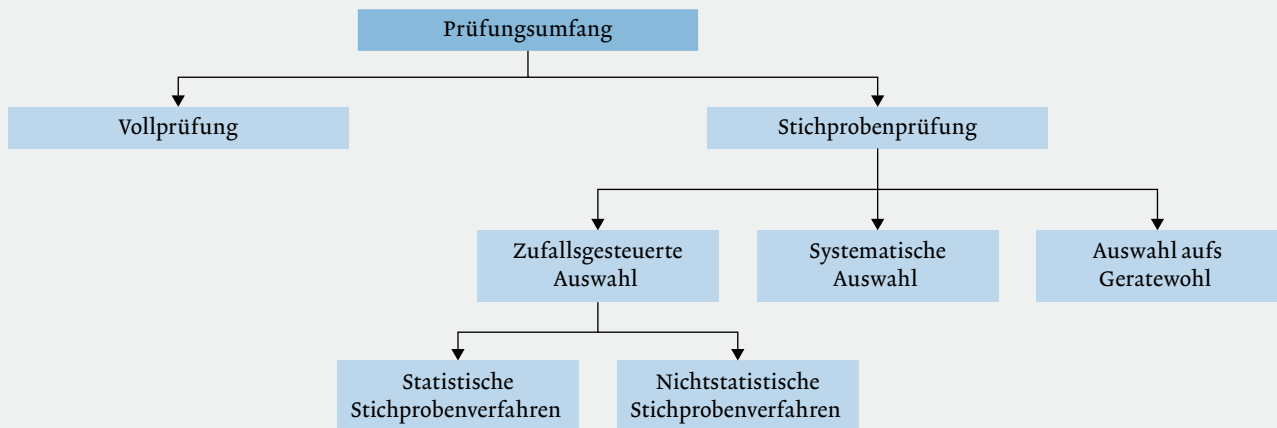


T. FLEMMING RUUD,  
 PROF. EM., PH.D., WP (N),  
 EMERITIERTER PROFESSOR  
 FÜR BETRIEBSWIRTSCHAFTSLEHRE (INTERNAL AUDIT, INTERNAL CONTROL),  
 UNIVERSITÄT ST. GALLEN



DAMIAN BORTH,  
 PROF. DR., LEHRSTUHL FÜR  
 KÜNSTLICHE INTELLIGENZ  
 UND MASCHINELLES  
 LERNEN, DIREKTOR AM  
 INSTITUTE OF COMPUTER  
 SCIENCE (ICS),  
 UNIVERSITÄT ST. GALLEN

Abbildung 2: **PRÜFUNGSUMFANG UND STICHPROBENAUSWAHLVERFAHREN** [10]



der Prüfer das Fehlerrisiko des Prüffelds. Gerade in der prüferischen Praxis stellt die Einschätzung des Fehlerrisikos oftmals eine Herausforderung dar. Dies gilt im Besonderen in Situationen der organisatorischen Veränderung eines Mandats, bspw. in Phasen der Restrukturierungen oder Unternehmensfusionen. Eine solche Herausforderung ergibt sich bspw. auch bei Erstprüfungen, in welchen Prüferinnen und Prüfer mit der Prüfung unbekannter Geschäftsprozesse betraut sind. In der praktischen Prüfungsdurchführung wissen Prüferinnen und Prüfer deshalb oftmals nicht, ob eine Stichprobe tatsächlich eine repräsentative Auswahl darstellt, welche die Fehlerrisiken adäquat adressiert [14]. Um diesen Herausforderungen zielgerichtet zu begegnen, wird nachfolgend ein KI-Verfahren vorgestellt, welches ermöglicht, repräsentative Stichproben aus umfangreichen Journalbuchungen zu erlernen.

**3. KÜNSTLICHE INTELLIGENZ, MASCHINELLES LERNEN UND DEEP LEARNING**

Im Allgemeinen hat die KI zum Ziel, kognitive Fähigkeiten menschlicher Intelligenz zu modellieren, um selbstlernend intellektuelle Aufgaben zu bewerkstelligen [15]. *Abbildung 3* zeigt eine beispielhafte Taxonomie der KI samt abgeleiteten Teildisziplinen, u. a. des Machine Learning (ML) [16]. Der Begriff ML bezeichnet ein Verfahren, das Computern die Fähig-

keit verleiht, selbstständig, d. h. ohne explizite menschliche Programmierung, Modelle für das Lösen einer vorgegebenen Aufgabenstellung zu erlernen [17]. Die Eigenschaft des selbstständigen Lernens beschreibt den Umstand, dass sich die Modellqualität durch die Zuführung von Informationen bzw. Daten im Zeitverlauf kontinuierlich verbessern kann [18]. In einem dem maschinellen Lernprozess vorgelagerten Schritt werden hierzu relevante Datenmerkmale manuell aus den Rohdaten extrahiert bzw. aufbereitet. Im Anschluss ist das ML-Verfahren angehalten, relevante Zusammenhänge bzw. Muster in den Merkmalen zu erkennen, um aus diesen ein gewünschtes Lösungsverhalten abzuleiten.

Eine noch «junge» Weiterentwicklung klassischer ML-Verfahren umfasst die Teildisziplin des Deep Learnings (DL). Im Allgemeinen bezeichnet DL den Einsatz selbstlernender künstlicher neuronaler Netze. Hierbei handelt es sich um ein biologisch inspiriertes Verfahren, das sich an Aufbau und Funktionsweise des menschlichen Gehirns orientiert, um ein intelligentes Verhalten zu imitieren. In Analogie zu den Neuronen des menschlichen Kortex bestehen DL-Verfahren aus einer Vielzahl von Schichten miteinander verknüpfter künstlicher Neuronen [20]. Die Eigenschaft «deep» bezieht sich auf die hohe Anzahl von Schichten künstlicher Neuronen. *Abbildung 4* zeigt eine schematische Darstellung eines künstlichen Neurons, bestehend aus Eingangsgewichten,

Abbildung 3: **BEISPIELHAFTER TAXONOMIE KÜNSTLICHER INTELLIGENZ** [19]

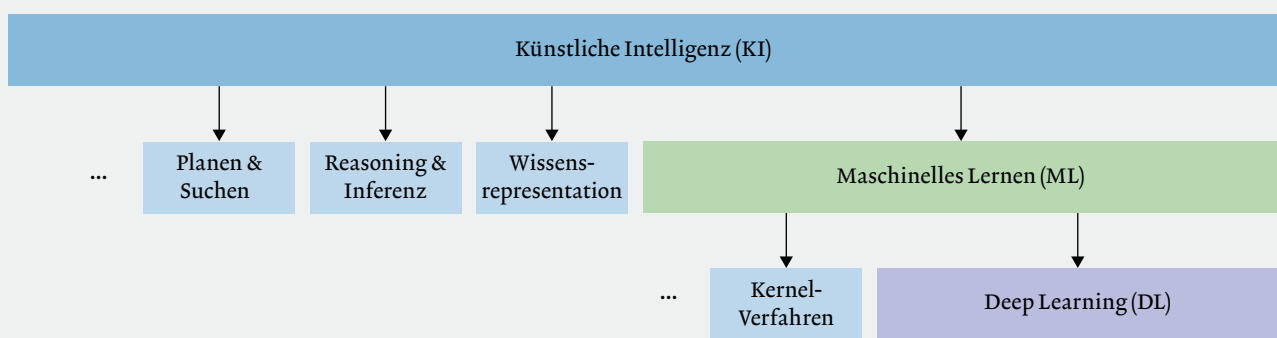
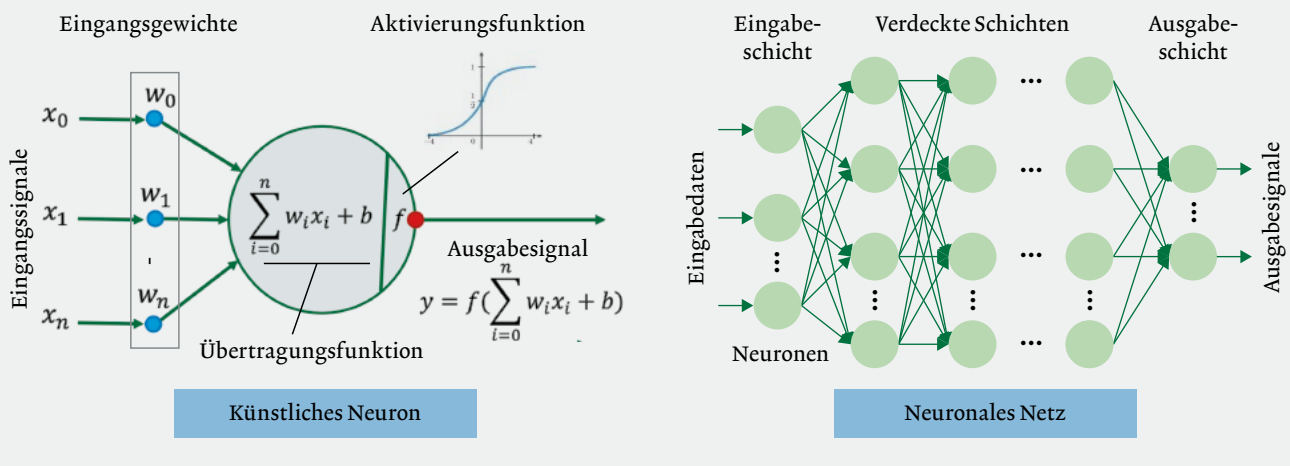


Abbildung 4: SCHEMATISCHE DARSTELLUNG EINES KÜNSTLICHEN NEURONS UND KÜNSTLICHEN NEURONALEN NETZES [22]



Aktivierungsfunktion und Ausgabesignal. *Abbildung 4* zeigt beispielhaft die grundlegende Struktur künstlicher neuronaler Netze, bestehend aus mehreren in Schichten angeordneten künstlichen Neuronen [21].

Das Lernen erfolgt anhand eines Trainingsprozesses, in welchem dem neuronalen Netz Trainingsdaten zugeführt und zugleich die gewünschte Lösung einer Aufgabenstellung aufgezeigt wird. Weicht die für ein Trainingsbeispiel beobachtbare Lösung von der gewünschten Lösung ab, erfolgt eine zielgerichtete Anpassung der Neuronengewichte. Dieser Trainingsprozess wird iterativ wiederholt, bis sich die beobachtbare und die gewünschte Lösung nicht mehr signifikant unterscheiden. Am Ende des Trainings verfügt das neuronale Netz über eine optimale Konfiguration der Gewichte, d. h. ein «Modell» für das Lösen der vorgegebenen Aufgabenstellung [23, 24].

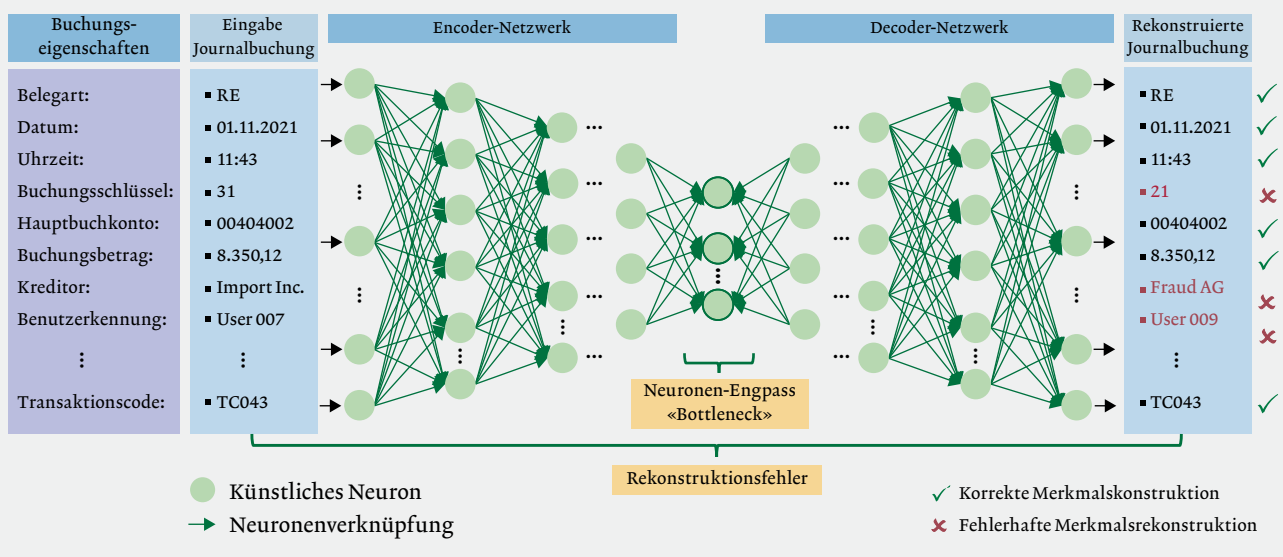
Im Unterschied zu den zuvor beschriebenen klassischen ML-Verfahren verfügen DL-Verfahren über die Fähigkeit, die

für das Lösen der Aufgabenstellung relevanten Datenmuster, sog. Repräsentationen, eigenständig aus den zugeführten Rohdaten zu erlernen. Ein solches gleichzeitiges Lernen relevanter Repräsentationen und abgeleiteten Lösungsverhaltenes für eine Aufgabenstellung wird als sogenanntes «Ende-zu-Ende-Lernen» bezeichnet [25]. Darüber hinaus ermöglichen DL-Verfahren ein hypothesenfreies «unüberwachtes» Lernen aus zugeführten Daten. Die Synthese beider Lernverfahren ist Ausgangspunkt für das Erlernen charakteristischer Repräsentation z. B. aus umfangreichen Journalbuchungen [26] und soll nachfolgend vorgestellt werden.

#### 4. KÜNSTLICHE-INTELLIGENZ-BASIERTE STICHPROBENAUSWAHL

Die innerhalb von ERP-Systemen erfassten Journalbuchungen sind oftmals gekennzeichnet durch eine umfangreiche Anzahl unterschiedlicher Merkmalsausprägungen (z. B. die Viel-

Abbildung 5: SCHEMATISCHER AUFBAU EINES TIEFEN AUTOENCODER-NETZES UND BEISPIELHAFT REKONSTRUKTION EINER JOURNALBUCHUNG [30]



zahl gebuchter Kreditoren, Hauptbuchkonten oder Belegarten) sowie eine umfangreiche Anzahl *unterschiedlicher Merkmalsbeziehungen* (z. B. die Vielfalt gebuchter Kombinationen aus Belegarten, Buchungsschlüsseln, Haupt- und Nebenbüchern). Ein für das unüberwachte «Ende-zu-Ende»-Lernen komplexer Merkmalsausprägungen und -beziehungen besonders geeignetes DL-Verfahren findet sich in sog. tiefen neuronalen Autoencoder-Netzen (im Folgenden als Autoencoder bezeichnet). Autoencoder bestehen aus zwei miteinander verknüpften neuronalen Netzen, die als Encoder- bzw. Decoder-Netzwerk bezeichnet werden [27]. Im Regelfall weisen beide Netze eine symmetrische Architektur auf und umfassen jeweils eine Vielzahl von Schichten künstlicher Neuronen [28].

*Abbildung 5* zeigt die schematische Architektur eines Autoencoders sowie die beispielhafte Rekonstruktion einer Journalbuchung. Die Zielsetzung des Autoencoder-Trainings besteht in der möglichst fehlerfreien Rekonstruktion der Merkmale zugeführter Journalbuchungen. Um zu verhindern, dass der Autoencoder lediglich lernt, die zugeführten Buchungsmerkmale von der Eingabeschicht zur Ausgabeschicht zu leiten, verknüpft ein Neuronen-Engpass (*engl.* Bottleneck) die Schnittstelle zwischen Encoder- und Decoder-Netzwerk. Die Modellierung eines solchen Engpasses erfolgt anhand der signifikanten Reduktion der Neuronen in den inneren Netzwerkschichten des Autoencoders [29].

Aufgrund der Engpass-Struktur ist der Autoencoder im Verlauf des Netzwerktrainings herausgefordert, komprimierte Repräsentationen der in den Journalbuchung enthaltenen Merkmalsausprägungen und -beziehungen zu lernen [31]. In jeder Schicht des Encoder-Netzwerks vollzieht sich eine nichtlineare Dimensionsreduktion der eingehenden Signale. In Analogie vollzieht sich in den Schichten des Decoders-Netzwerks ein Prozess der nichtlinearen Dimensions-Rekonstruktion. Mit fortschreitendem Training gelingt es dem Autoencoder kontinuierlich, charakteristischere Repräsentationen der zugeführten Journalbuchungen zu erlernen. Die Repräsentationen ermöglichen die zunehmend fehlerfreie Rekonstruktion einer Vielzahl der zugeführten Buchungen, d. h., der Rekonstruktionsfehler des Modells wird während des Lernprozesses fortlaufend minimiert.

*Abbildung 6* illustriert das Lernen charakteristischer Buchungseigenschaften durch den Autoencoder anhand der Rekonstruktion zugeführter Buchungen. Ausgehend von der Grundgesamtheit an Journalbuchungen verarbeitet die Eingabeschicht des Encoders zunächst die originären Merkmale der zugeführten Buchungen. Bedingt durch den Neuronen-Engpass lernt der Encoder innerhalb der nachfolgenden Schichten zunehmend charakteristischere und zugleich komprimiertere Repräsentationen. In den frühen Schichten entsprechen die Repräsentationen zunächst einfachen Merkmalsausprägungen und -beziehungen. In den tieferen Schichten stellen die erlernten Repräsentationen die in den Journalbuchungen enthaltenen komplexen Buchungssystematiken dar.

Eine im Kontext der Stichprobenauswahl bedeutsame Autoencoder-Variante stellen sog. «Vector Quantised-Variational Autoencoder» (VQ-VAE) dar [32]. Das Lernziel der Vektor-Quantisierung (*engl.* Vector-Quantisation) besteht in der

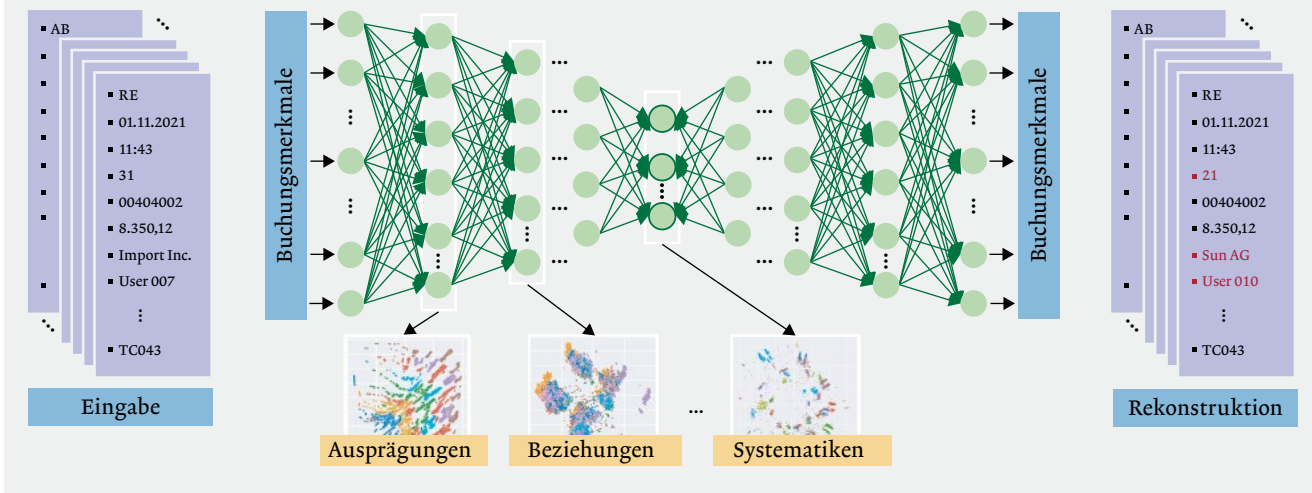
Rekonstruktion zugeführter Daten auf der Grundlage einer möglichst geringen Anzahl charakteristischer Repräsentationen sog. Quantisierungen. Die Übertragung dieser Idee auf den Prüfungskontext bedeutet, dass die Quantisierungen als repräsentative Stichprobenauswahl der Trainingsdaten interpretiert werden können. Im Folgenden werden die erlernten Quantisierungen deshalb als eine durch den VQ-VAE erlernte Stichprobenauswahl betrachtet.

## 5. PRAKTISCHE ANWENDUNG IN DER PRÜFUNGSPRAXIS

Die praktische Anwendung des VQ-VAE-Verfahrens wurde anhand einer explorativen Studie untersucht [33]. Aufgrund der oftmals hohen Vertraulichkeit realer Journalbuchungen basiert die Studie auf öffentlich zugänglichen Daten. So wurde das Verfahren u. a. auf einem Datensatz kreditorischer Zahlungen der Stadt Philadelphia [34] evaluiert. Der Datensatz umfasst 238 894 Zahlungsbuchungen der über 60 städtischen Departments und Ämter für das Jahr 2017, die ein Gesamtvolumen i. H. v. USD 4,2 Mrd. aufweisen. Die Zahlungen weisen eine hohe technische und semantische Ähnlichkeit zu innerhalb von ERP-System erfassten Zahlungen auf, bspw. der in SAP FI durch den Transaktionscode F110 gesteuerten Zahläufe. Der trainierte VQ-VAE umfasste 22 Netzschichten bestehend aus insgesamt 18 378 künstlichen Neuronen. Um eine visuelle Interpretation der erlernten Stichprobenauswahl durch Prüferinnen und Prüfer zu ermöglichen, umfasst die mittlere Schicht des Neuronen-Engpasses zwei Neuronen.

*Abbildung 7* illustriert beispielhaft die erlernte Stichprobenauswahl innerhalb des zweidimensionalen Raums des Neuronen-Engpasses für einen jeweils unterschiedlichen Stichprobenumfang  $n=8, 32$  und  $64$ . Die farbig dargestellten Bildpunkte entsprechen den erlernten Repräsentationen einzelner Zahlungen. Die rot hervorgehobenen und zugleich nummerierten Bildpunkte visualisieren die erlernte Stichprobenauswahl. Im Rahmen des Lernprozesses wird jeder erlernten Zahlungsrepräsentation genau eine Stichprobe zugeordnet (vgl. *Abbildung 7*, exemplarisch dargestellte Kreise). Die einer Stichprobe zugeordneten Zahlungsrepräsentationen weisen

Abbildung 6: **BEISPIELHAFTES LERNEN CHARAKTERISTISCHER BUCHUNGSREPRÄSENTATIONEN**



eine identische Farbkodierung auf (vgl. *Abbildung 7*, exemplarisch pro Kreis eingefärbte Zahlungsrepräsentationen). Auf der Grundlage der erlernten Stichprobenauswahl ist es dem VQ-VAE möglich, die Mehrheit der ursprünglichen Zahlungsbuchungen zu rekonstruieren.

Wie unscharf die Rekonstruktion einer Journalbuchung auf Grundlage einer erlernten Stichprobe erfolgen kann, wird anhand des Rekonstruktionsfehlers gemessen. Der Fehler stellt zugleich ein Gütekriterium der Stichprobenauswahl dar und bestimmt den Grad der Repräsentanz einer Stichprobe für die jeweils zugeordneten Zahlungsrepräsentationen. Je grösser die Distanz zwischen Stichprobe und zugeordneter Zahlungsrepräsentation, umso unschärfer erfolgt die Rekonstruktion der ursprünglichen Zahlungsbuchung. Für eine gelernte Stichprobenauswahl bedeutet dies, dass der Grad der Repräsentanz anhand des Rekonstruktionsfehlers gemessen und in Abhängigkeit des prüferischen Ermessens variiert werden kann. Die Erhöhung des Stichprobenumfangs führt dazu, dass die erlernte Stichprobenauswahl zunehmend detaillierte

Merkmalsausprägungen und -eigenschaften der Zahlungsbuchungen repräsentiert (vgl. *Abbildung 7*, Stichprobe n=32 und 64). Die räumliche Anordnung der Zahlungsrepräsentationen im zweidimensionalen Raum spiegelt zudem die semantische Ähnlichkeit der verschiedenen Zahlungsbuchung wider. Semantisch ähnliche Zahlungen sind durch eine geringe räumliche Nähe bzw. Dichte gekennzeichnet; unähnliche Zahlungen durch eine entsprechende hohe Distanz.

*Abbildung 8* zeigt das Ergebnis einer erlernten Stichprobenauswahl des Umfangs n=128 der insgesamt 238894 Zahlungen. In *Abbildung 8* sind beispielhaft die erlernten repräsentativen Stichproben Nr. 16 und 99 dargestellt samt den jeweils repräsentierten Zahlungsbuchungen. Die Detailanalyse zeigt, dass die Stichprobe Nr. 16 Zahlungen repräsentiert, die semantisch den Einkauf von Transportdienstleistungen betreffen, welche durch American Express Travel Vouchers beglichen wurden. Stichprobe Nr. 99 hingegen repräsentiert Zahlungen des Departments 25 Fleet Management, welche semantisch den Einkauf von Fahrzeugteilen und -zubehör umfassen.

Abbildung 7: **ERLERNTE STICHPROBENAUSWAHL BEI UNTERSCHIEDLICHEM STICHPROBENUMFANG** [35]

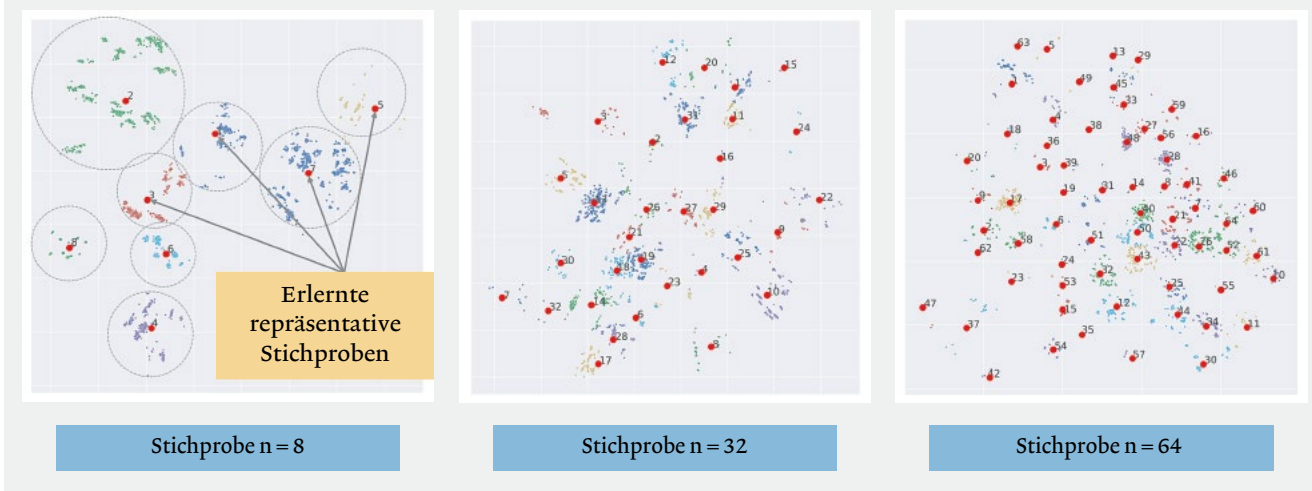


Abbildung 8: **ERLERNTE STICHPROBENAUSWAHL UND BEISPIELE REPRÄSENTIERTER ZAHLUNGEN** [36]



Stichprobe n = 128

Erlernte repräsentative Stichprobe (99)								
FY	Datum	Zahlung	Department	Typ	Object	Lieferant	Details	Betrag USD
2017	20.04.17	CHEK17121800	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	L. SAYRE INCORPORATED	auto parts	132.59

Repräsentierte Philadelphia City Payments								
FY	Datum	Zahlung	Department	Typ	Object	Lieferant	Details	Betrag USD
2017	14.07.16	CHEK17005079	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	TRANSTECK INC	auto parts	-32.17
2017	20.04.17	CHEK17121800	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	G.L. SAYRE INCORPORATED	auto parts	132.59
2017	04.05.17	CHEK17125451	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	TRANSTECK INC	auto parts	37.60
2017	01.06.17	CHEK17136892	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	CHAPMAN FORD SALES	auto parts	-293.71
2017	01.06.17	CHEK17138737	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	JACK DOHERTY COMPANIES INC	auto parts	539.70
2017	08.06.17	CHEK17139311	25 FLEET MANAGEMENT	03 MATERIALS AND SUPPLIES	MOTOR VEHICLE PARTS AND ACCESSORIES 0328	TRANSTECK INC	auto parts	11.46
...	...	...	...	...	...	...	...	...

Erlernte repräsentative Stichprobe (16)								
FY	Datum	Zahlung	Department	Typ	Object	Lieferant	Details	Betrag USD
2017	23.06.17	ACHD17189704	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	363.2

Repräsentierte Philadelphia City Payments								
FY	Datum	Zahlung	Department	Typ	Object	Lieferant	Details	Betrag USD
2017	22.07.16	ACHD17011732	13 FIRE	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	228.68
2017	23.06.17	ACHD17189704	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	113.00
2017	19.08.16	ACHD17026108	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	354.20
2017	28.10.16	ACHD17059715	22 HUMAN SERVICES	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	539.20
2017	23.12.16	ACHD17090986	42 COMMERCE	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	213.00
2017	17.02.17	ACHD17121338	22 HUMAN SERVICES	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	American Express travel vouchers	320.60
...	...	...	...	...	...	...	...	...

**6. FAZIT UND AUSBLICK**

Zusammenfassend lässt sich konstatieren, dass VQ-VAE-basierte Verfahren Prüferinnen und Prüfer dabei unterstützen können, repräsentative Stichproben automatisiert zu erlernen. Dies gilt im Besonderen für Situation in fehlender bzw. unvollständiger Vorinformation über ein Prüffeld. Zugleich entfällt das mit der statistischen Stichprobenauswahl assoziierte Risiko der Auswahl ungeeigneter Parametrisierungen deskriptiver Statistiken, bspw. durch die Annahme unpassender statistischer Verteilungen. Von besonderer Bedeutung ist für die Praxis hierbei, dass der Grad der Repräsentanz der gelernten Stichprobenauswahl exakt durch den Rekonstruktionsfehler bestimmt werden kann. In Analogie zu statistischen Auswahlverfahren, ist es hierdurch möglich, Stichprobenrisiken durch die Anpassung des Stichprobenumfangs zielgerichtet zu variieren (vgl. *Abbildung 7*). Welcher Stichprobenumfang für die Grundgesamtheit ausgewählt wird, liegt nach wie vor im professionellen Ermessen der Prüferin bzw. des Prüfers.

Ein weiteres Alleinstellungsmerkmal von VQ-VAE-basierten Verfahren im prüferischen Kontext ist das Erlernen von erklärbaren Repräsentationen der Journalbuchungen (vgl. *Abbildung 8*). Bedingt durch die Dimensionsreduktion ist es Prüferinnen und Prüfern bspw. visuell möglich, die Grundge-

samtheit eines Prüffelds ganzheitlich zu erfassen und die Homogenität (hohe Dichte) und etwaige Abweichungen (geringe Dichte) zu beurteilen. Im Kontext nichtstatistischer Auswahlverfahren kann eine solche Darstellung Ausgangspunkt einer zielgerichteten Stichprobenauswahl sein. Dieses Vorgehen ermöglicht es Prüferinnen und Prüfern, Fehler- und Fraud-Risiken zielgerichtet zu adressieren.

Die stichprobenbasierte Prüfung spielt eine bedeutende Rolle im Rahmen der wirtschaftlichen Durchführung von Prüfungshandlungen [37]. Ob es sich bei dem vorgestellten Verfahren um eine zufallsgesteuerte, systematische oder gar eigene Kategorie der Stichprobenauswahl handelt, bedarf aufgrund der Neuartigkeit von KI-unterstützten Auswahlverfahren der Diskussion.

Das vorgestellte Verfahren zeigt erneut, dass sich die prüferische Praxis in einer Transformation von der IT-gestützten Prüfung hin zu einer zukünftig zunehmend durch KI-unterstützten Prüfung befindet [38]. Die Anwendung von KI im Kontext der repräsentativen Stichprobenauswahl bzw. des Stichprobenlernens stellt einen weiteren Schritt in eine solche Zukunft dar. Die Implementierung in der Praxis wird zeigen, inwieweit ein solches Verfahren einen nutzenstiftenden Beitrag für die Prüfung leisten kann [39].

**Fussnoten:** \*Die Autoren bedanken sich bei Monika Heyder, Irene Lüthy und Marcel Meier für die zahlreichen, wertvollen Anregungen und die kritische Durchsicht des Manuskripts.

- 1)** International Accounting Standard (ISA) 1, Presentation of Financial Statements, International Federation of Accountants (IFAC), 2007. **2)** Expert-suisse, Handbuch der Wirtschaftsprüfung, Band Ordentliche Revision, 2016. **3)** S. V. Grabski, S. A. Leech und P. J. Schmidt, A Review of ERP Research: A Future Agenda for Accounting Information Systems, *Journal of Information Systems*, 25(1):37–78, 2011. **4)** M. Schreyer, T. Sattarov, D. Borth, A. Dengel und B. Reimer, Detection of Anomalies in Large Scale Accounting Data using Deep Autoencoder Networks, arXiv preprint arXiv:1709.05254, 2017. **5)** M. Schreyer, T. Sattarov, C. Schulze, B. Reimer und D. Borth, Detection of Accounting Anomalies in the Latent Space using Adversarial Autoencoder Neural Networks, 2nd KDD Workshop on Anomaly Detection in Finance, Anchorage, Alaska, USA, 2019. **6)** M. Schultz and M. Tropmann-Frick, Autoencoder Neural Networks Versus External Auditors: Detecting Unusual Journal Entries in Financial Statement Audits, In Proceedings of the 53<sup>rd</sup> Hawaii International Conference on System Sciences, 2020. **7)** I. Bhattacharya and E. Roos Lindgreen, A Semi-Supervised Machine Learning Approach to Detect Anomalies in Big Accounting Data, In Proceedings of the European Conference on Information Systems (ECIS), 2020. **8)** J. Nonnenmacher, F. Kruse, G. Schumann und J. Marx Gómez, Using Autoencoders for Data-Driven Analysis in Internal Auditing, In Proceedings of the 54<sup>th</sup> Hawaii International Conference on System Sciences (p. 5748), 2021. **9)** Siehe Fn. 2. **10)** Siehe Fn. 2. **11)** G. Jokovich, Statistical Sampling in Auditing, *International Journal of Accounting and Financial Management*, 16:892–898, 2013. **12)** D. M. Guy, D. R. Carmichel und O. R. Whittington, Audit Sampling – An Introduction, John Wiley & Sons, 2002. **13)** Siehe Fn. 2. **14)** T. F. Ruud, K. Schramm und A. Allgaier, Leitlinie zum Internen Audit, Institute of Internal Auditing (IIA) Switzerland, 4. Auflage, 2021. **15)** S. Russell und P. Norvig, Artificial Intelligence: A Modern Approach, Third Edition, Pearson Education Ltd., pp. 1–5, 2016. **16)** D. Borth, Introduction to Machine Learning and Deep Learning, Lehrveranstaltung an der Universität St. Gallen (HSG), 2020. **17)** A. L. Samuel, Some Studies in Machine Learning Using the Game of Checkers, *IBM Journal of Research and Development*, 3(3), pp. 210–229, 1959. **18)** C. M. Bishop, Pattern Recognition and Machine Learning, Springer Science LLC, pp. 3–4, 2006. **19)** siehe Fn. 16. **20)** Y. LeCun, Y. Bengio, and G. Hinton, Deep Learning, *Nature*, 521(7553), pp. 436–444, 2015. **21)** Y. Bengio, A. Courville und P. Vincent, Representation Learning: A Review and New Perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp. 1798–1828, 2013. **22)** Siehe Fn. 20. **23)** Siehe Fn. 18. **24)** Siehe Fn. 20. **25)** Siehe Fn. 21. **26)** Siehe Fn. 5. **27)** G. E. Hinton und R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, *Science*, 313(5786), pp. 504–507, 2006. **28)** S. Hawkins, H. He, G. Williams und R. Baxter, Outlier Detection Using Replicator Neural Networks, In International Conference on Data Warehousing and Knowledge Discovery, pp. 170–180, Springer, Berlin, Heidelberg, 2002. **29)** Siehe Fn. 28. **30)** Siehe Fn. 4. **31)** Siehe Fn. 4. **32)** A. van den Oord, O. Vinyals und K. Kavukcuoglu, Neural Discrete Representation Learning, In Proceedings of the 31<sup>st</sup> International Conference on Neural Information Processing Systems, pp. 6309–6318, 2017. **33)** M. Schreyer, T. Sattarov, A. S. Gierbl, B. Reimer und D. Borth, Learning Sampling in Financial Statement Audits using Vector Quantised Variational Autoencoder Neural Networks, In Proceedings of the International Conference on Artificial Intelligence (ICAIF) 20, Association of Computing Machinery (ACM), 2020. Die vollständigen Studienergebnisse sind über die Digital Library der ACM durch die nachfolgende URL abrufbar (eingesehen am 28.12.2021): <https://dl.acm.org/doi/abs/10.1145/3383455.3422546>. **34)** Der vollständige Datensatz ist über das Datenportal der Stadt Philadelphia durch die nachfolgende URL abrufbar (eingesehen am 28.12.2012): <https://www.phila.gov/2019-03-29-philadelphias-initial-release-of-city-payments-data/> **35)** Siehe Fn. 33. **36)** Siehe Fn. 33. **37)** Siehe Fn. 2. **38)** T. Sun, Applying Deep Learning to Audit Procedures: An Illustrative Framework, *Accounting Horizons*, 33(3), pp. 89–109, 2019. **39)** S. Cho, M. A. Vasarhelyi, T. Sun, und C. Zhang, Learning from Machine Learning in Accounting and Assurance, *Journal of Emerging Technologies in Accounting*, 17(1), pp. 1–10, 2020.