MARCO SCHREYER
ANITA GIERBL
T. FLEMMING RUUD
DAMIAN BORTH

ARTIFICIAL INTELLIGENCE ENABLED AUDIT SAMPLING

Learning to draw representative and interpretable audit samples from large-scale journal entry data

Artificial Intelligence (AI) is increasingly perceived as a valuable technology in internal and external auditing. The following article introduces the application of deep learning (DL), a thriving subdiscipline of AI, to draw a learning-based representative audit sample from extensive volumes of financial accounting data.

1. SAMPLE-BASED AUDIT PROCEDURES

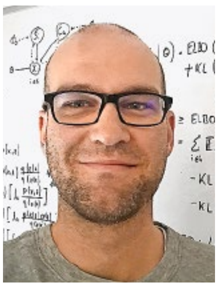
Financial statements serve as a fundamental basis [1] in today's economic decision-making by a wide range of interest groups and their trustworthiness is considered an essential pillar of good corporate governance. The audit of financial statements is designed to "obtain reasonable assurance about whether the financial statements as a whole are free from material misstatement" (ISA 200.5). Throughout an audit, auditors form an opinion on whether a statement is prepared, in all material respects, in accordance with the applicable financial reporting framework (ISA 700). To detect a material misstatement, the *International Standards on Auditing* (ISA) require auditors to assess a financial statement's underlying accounting transactions, using a process referred to as journal entries testing (ISA 240, ISA 500). *Figure 1* depicts a hierarchical view of the journal entry recording process in designated database tables of an ERP system. Organisations generate vast numbers of journal entries in Enterprise Resource Planning (ERP) systems [2]. For economic reasons, it is seldom possible to conduct a detailed test of the entire population of entries [3]. To efficiently audit large amounts of journal entries, auditors regularly conduct sample-based assessments using a technique referred to as *audit sampling*. This is particularly the case for "data-intensive" financial statement line items, e.g. revenues, accounts receivables, purchases and accounts payables.

Driven by rapid technological advances in AI, *deep learning* (DL) assisted audit techniques have entered audit practice [4, 5, 6, 7, 8]. These developments raise the following questions: Can DL also be used to learn to draw representative audit samples? And if so, can such sampling be learnt in a way that it remains interpretable by auditors? This article addresses both questions and introduces the application of a novel AI technique referred to as *Vector Quantised-Variational Autoencoders (VQ-VAE)*. Results of an empirical study illustrate the VQ-VAE's ability to assist auditors in the learning of representative and interpretable audit samples.

2. AUDIT SAMPLING AND SAMPLING RISK

Auditors use different sampling techniques to obtain audit evidence when performing audit procedures (ISA 530). *Audit sampling* is defined as the "application of audit procedures to less than 100% of items within a population" (ISA 530.5*a*). While sampling increases the efficiency of a financial audit, it also exhibits a *sampling risk*. The sampling risk denotes the likelihood "that the auditor's conclusion based on a sample may be different from the conclusion if the entire population were subjected to the same audit procedure" (ISA 530.5*c*). To mitigate sampling risks, the auditors aim to determine a sample that is representative of the entire journal entry population of an audit field [9], meaning that the sampled items reflect the structural characteristics of the audit population and are free of systematic bias [10].

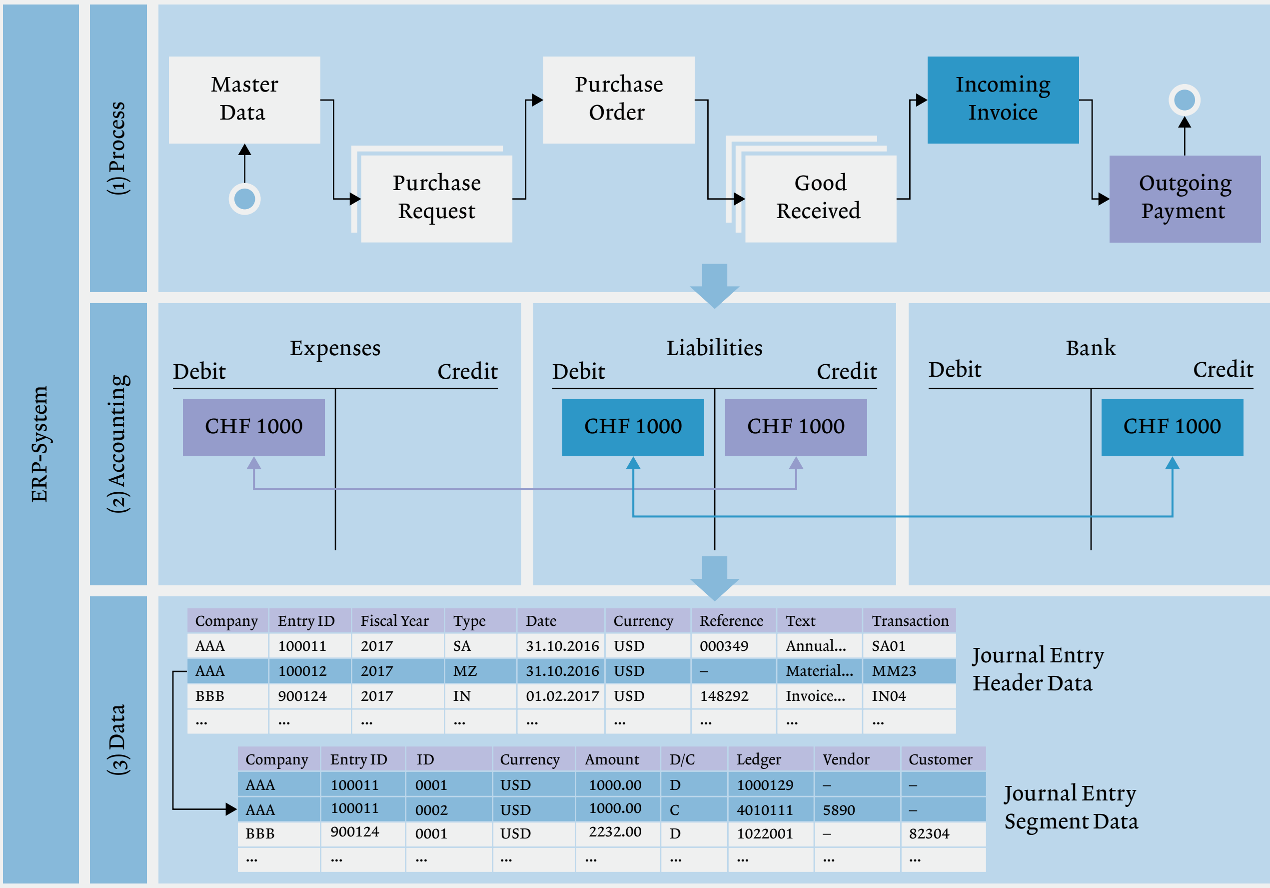
According to the ISA 530, *random, systematic,* and *haphazard* audit sampling techniques are distinguished (see *Figure 2*).



MARCO SCHREYER,
RESEARCHER, CHAIR OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, INSTITUTE OF COMPUTER SCIENCE (ICS), UNIVERSITY OF ST. GALLEN

ANITA GIERBL,
DR. OEC. HSG, RESEARCHER, UNIVERSITY OF ST. GALLEN, TECHNICAL ASSISTANT SWISS GAAP FER, AUDIT, PWC SWITZERLAND

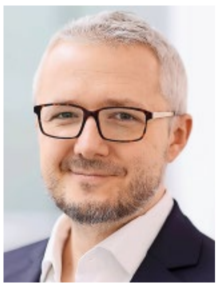
Figure 1: **SCHEMATIC ILLUSTRATION OF AN ENTERPRISE RESOURCE PLANNING (ERP) SYSTEM THAT RECORDS JOURNAL ENTRY INFORMATION AT DISTINCT LEVELS OF ABSTRACTION, NAMELY (1) THE BUSINESS PROCESS LEVEL, (2) THE ACCOUNTING LEVEL AND (3) THE DATA LEVEL**



Random sampling is classified into statistical and non-statistical sampling methods. In statistical sampling techniques samples are drawn based on objectifiable probabilities [12]. In non-statistical sampling techniques, a sample is drawn dependent on the auditor's professional judgment. *Systematic* sampling techniques are characterised by predefined sampling intervals [13]. When conducting *haphazard* sampling, the auditor selects the sample without following a structured technique.

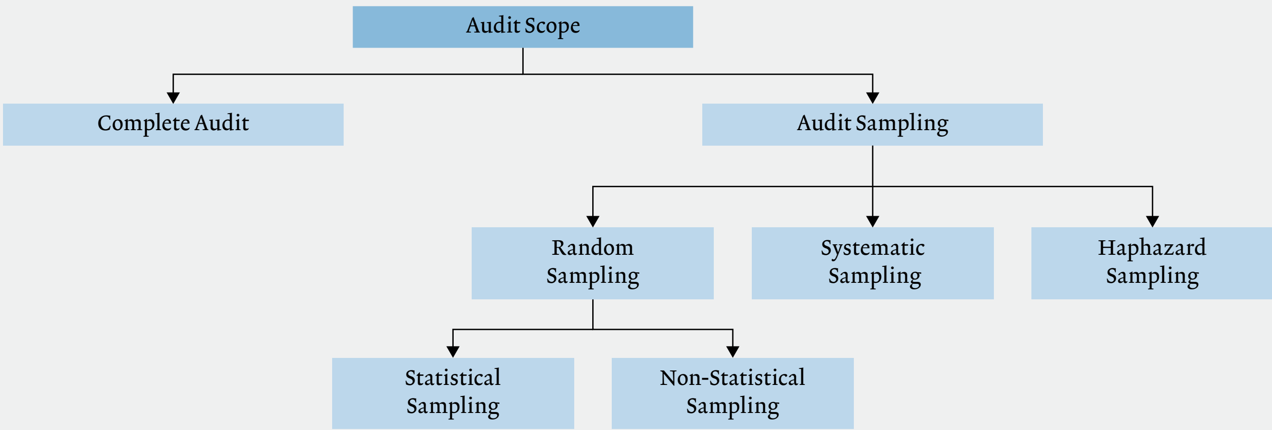
Auditors are responsible for determining which sampling procedure should be used to obtain sufficient audit evidence [14]. To choose a suitable sampling procedure, auditors are required to assess the characteristics of the audit population, e.g. in terms of homogeneity and structural characteristics. Throughout the assessment, the auditor evaluates the risk of material misstatement. In practice, such assessments can be challenging, particularly when the organisation subject to the audit is experiencing major changes, e.g. in phases of organisational restructuring or corporate mergers. Difficulties may also arise in new audit engagements where auditors are not familiar with the client's business processes and associated risks. In both situations, auditors might be uncertain if the sample corresponds to a representative selection reducing the sampling risk sufficiently [15]. In the following chapter, an AI method is presented that learns to draw representative audit samples from large volumes of journal entry data.

T. FLEMMING RUUD,
PROF. EM., PH.D., CPA (NO),
CHAIR OF BUSINESS
ADMINISTRATION (INTERNAL
AUDIT, INTERNAL CONTROL),
UNIVERSITY OF ST. GALLEN



DAMIAN BORTH,
PROF. DR., CHAIR OF
ARTIFICIAL INTELLIGENCE
AND MACHINE LEARNING,
DIRECTOR AT THE
INSTITUTE OF COMPUTER
SCIENCE (ICS),
UNIVERSITY OF ST. GALLEN

Figure 2: **OVERVIEW OF AUDIT SCOPE AND AUDIT SAMPLING PROCEDURES** [11]



3. ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND DEEP LEARNING

AI aims to model the cognitive abilities of humans to accomplish intellectual tasks through learning [16]. *Figure 3* shows a taxonomy of AI and its derived subdisciplines, including *machine learning* (ML) [17]. ML refers to a process that enables computers to learn models for solving a given task *"without being explicitly programmed"* [18]. The "learning" characteristic entails that the model quality can continuously improve over time when processing new data [19]. To facilitate such learning, domain specific data features need to be "manually" defined and extracted from the raw data. Subsequently, ML models are trained to solve a predefined task by learning relevant patterns from the manually extracted data features.

This traditional ML setup is enhanced by *deep learning* (DL), a subdiscipline of ML. In general, DL refers to a neuroscience-inspired approach, imitating the function of a biological brain. In analogy to the neurons of a human brain, DL methods encompass a collection of *artificial neurons*. *Figure 4* illustrates an individual artificial neuron consisting of input weights, activation function, and output signal. To conceptually simulate the function of the brain, the distinct neurons are interconnected in an artificial neural network. *Figure 4* shows an example of a basic artificial neural networks structure, consisting of multiple artificial neurons arranged in layers [21]. The "deep" characteristic refers to the often extensive number of artificial neuron layers [22], i. e. the depth of the neural network.

During network training raw data are continuously fed to the neural network. Based on the input data the neural network attempts to create an output that is a solution to a predefined task. In situations where a prediction proposed by the network deviates from the expected solution, the neuron weights of the network are adjusted. This training process is repeated until the proposed and expected solutions converge. At the end of the training procedure, the neural network encapsulates an optimal weight configuration, i. e. a *model,* able to solve the task [24, 25].

In contrast to traditional ML techniques, DL exhibits the ability to directly learn data features, also referred to as *representations,* relevant to solving a given task from the raw data [26]. The parallel "end-to-end" learning of (i) data representations and (ii) task solution defines the foundation for learning-based representative audit sampling [27].

4. ARTIFICIAL INTELLIGENCE BASED AUDIT-SAMPLING

Journal entries recorded in ERP systems are characterised by a broad number of different *accounting attribute characteristics* (e. g. the multitude of posted vendors, general ledger accounts, document types) and often exhibit a variety of distinct *account-*

Figure 3: **EXAMPLE TAXONOMY OF ARTIFICIAL INTELLIGENCE (AI)** [20]

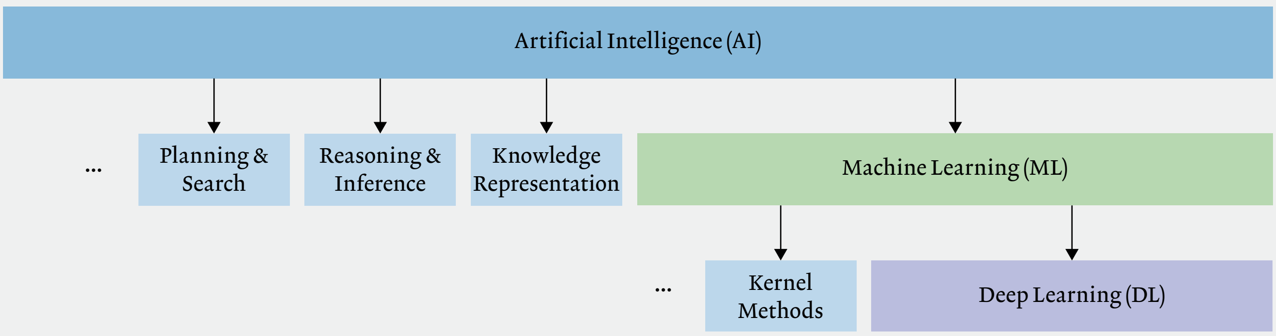
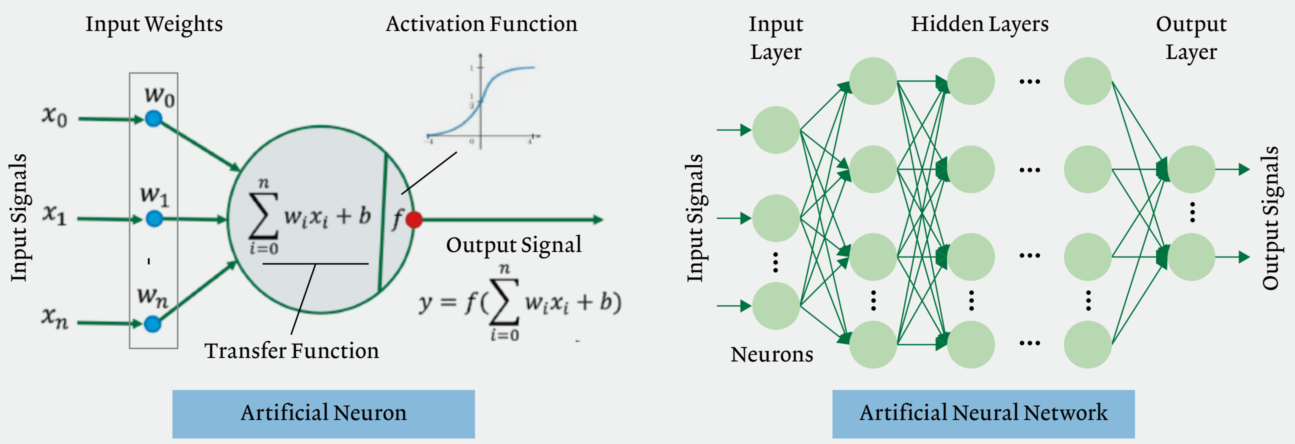


Figure 4: **SCHEMATIC ILLUSTRATION OF AN ARTIFICIAL NEURON AND ARTIFICIAL NEURAL NETWORK** [23]



ing attribute relationships (e.g. the variety of posted combinations of document types, posting keys, general ledgers, and subledgers). A DL method particularly suited to the *end-to-end* learning of complex accounting data can be found in *autoencoder neural networks* (hereafter referred to as *autoencoders*). Autoencoders encompass two interconnected neural networks, referred to as *encoder* and *decoder networks,* respectively [28]. Usually the two networks exhibit a symmetric structure, and are comprised of several layers of artificial neurons [29].

Figure 5 illustrates the general architecture of the way an autoencoder learns representations of journal entry data. The objective of the autoencoder training is to reconstruct the attributes of a given journal entry as faithfully as possible. To prevent the autoencoder from merely learning to forward the entry attributes from the input layer to the output layer, the number of neurons of the inner network layers is reduced, creating an information bottleneck. The bottleneck establishes the interconnection between the encoder and the decoder network [30].

Due to the bottleneck structure, each encoder layer applies a non-linear dimensionality reduction to the input journal entry attributes and the encoder learns a compressed representation of each individual journal entry. Subsequently, each decoder layer conducts a non-linear dimensionality reconstruction of the journal entry representation. The difference between a reconstructed journal entry and the original input is measured as *reconstruction error.* As training progresses, the autoencoder continuously minimises the reconstruction error. Hence, the error implicitly determines to which extent the journal entry characteristics and relationships are accumulated by the learned low-dimensional representations.

Figure 6 illustrates a typical autoencoder learning process of journal entry representations. Initially, the input layers process the accounting attributes of an entry. In the subse-

Figure 5: **ARCHITECTURE OF A DEEP AUTOENCODER NETWORK AND EXAMPLE JOURNAL ENTRY RECONSTRUCTION** [31]

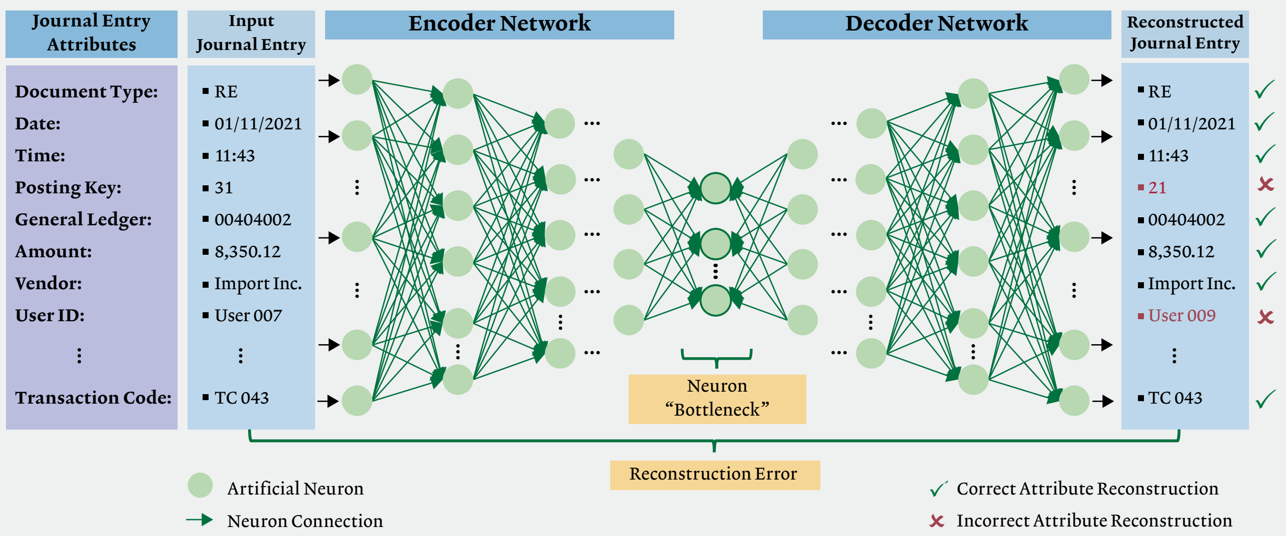
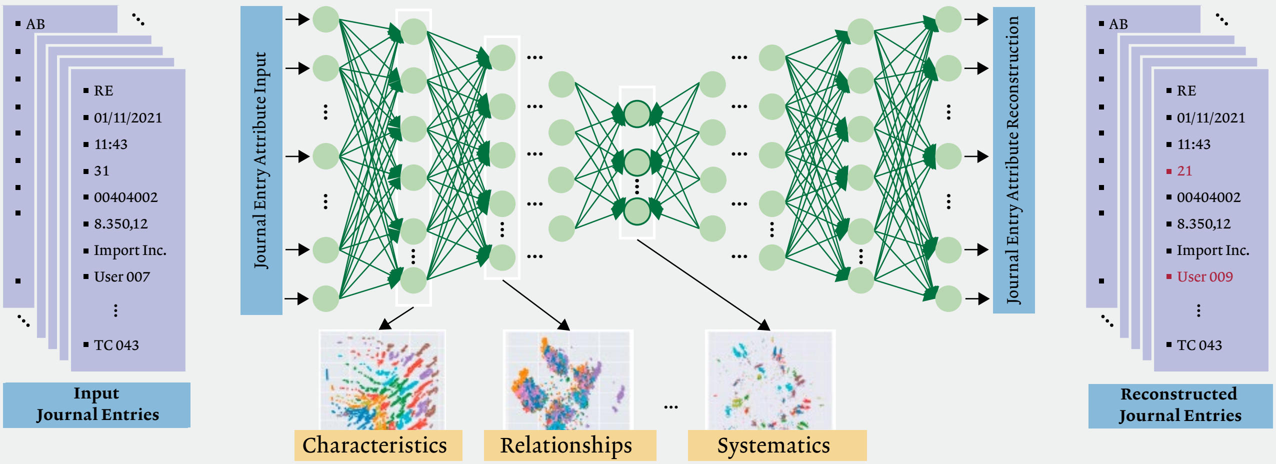


Figure 6: **LEARNING SCHEME OF JOURNAL ENTRY REPRESENTATIONS AT DIFFERENT LEVELS OF ACCOUNTING ABSTRACTION**



quent layers, the encoder extracts representations that correspond to an increased level of accounting abstraction. In the initial layers, the learned representations reflect simple accounting attribute characteristics and relationships. With subsequent layers, the learned representations reflect the complex accounting systematics evident in the journal entries.

In the context of audit sampling, an effective autoencoder variant is the *Vector Quantised-Variational Autoencoder* (VQ-VAE) [32]. VQ-VAEs denote a specialized autoencoder with the learning objective to reconstruct the majority of journal entries based on a minimal number of learned representations.

5. APPLICATION IN AUDIT PRACTICE

The practical application of VQ-VAEs for the purpose of audit sampling was investigated by the authors in an empirical study [33]. Due to the high confidentiality of real-world journal entry data, the study is based on publicly available datasets. One dataset used in the study encompasses 238,894 vendor payments by the City of Philadelphia recorded in 2017 [34]. The payments, totalling USD 4.2 billion, originate from 60 city departments and agencies, and exhibit a high technical and semantic similarity to periodic payment runs usually recorded in SAP ERP systems, e. g. executed via transaction code F110. The trained VQ-VAE encompasses 22 network layers including 18,378 artificial neurons. The VQ-VAE bottleneck layer comprises two neurons to allow for a visual interpretation of the learned audit sample.

Figure 7 illustrates the learning-based audit samples within the two-dimensional space of the neuron bottleneck for different audit sample sizes n = 8, 32, and 64. The coloured dots depict the learned representations of the entire population of city payments. The larger red dots indicate the learning-based audit sample. Each learned payment representation (smaller coloured dots) corresponds to precisely a single learned audit sample (larger red dot). The red circles indicate distinct sets of payments (exhibiting an identical colour coding) represented by the same audit sample respectively. Based on the learned audit samples, the VQ-VAE exhibits the

Figure 7: **LEARNED REPRESENTATIVE AUDIT SAMPLES OF DIFFERENT SAMPLE SIZES** [35]

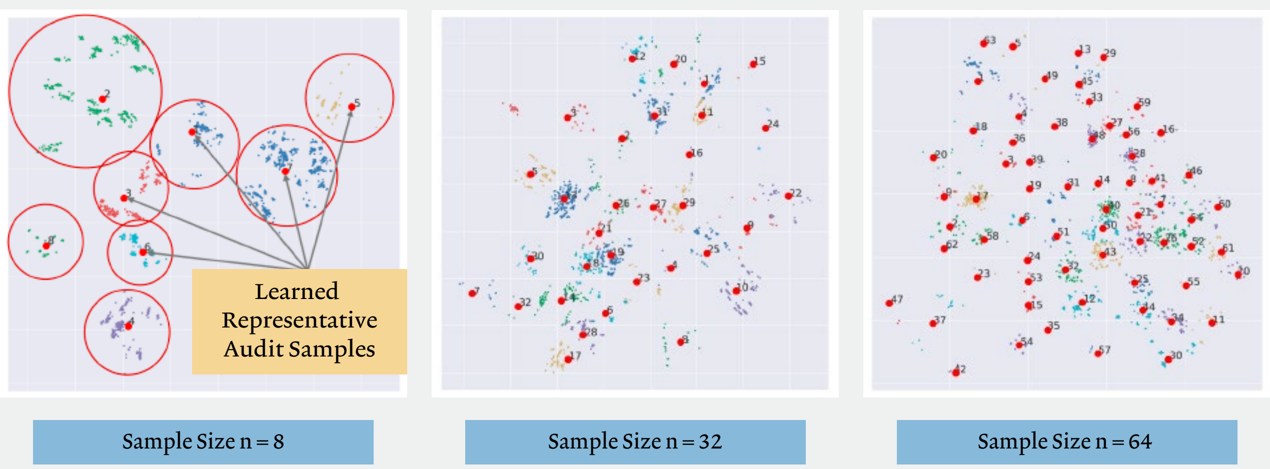
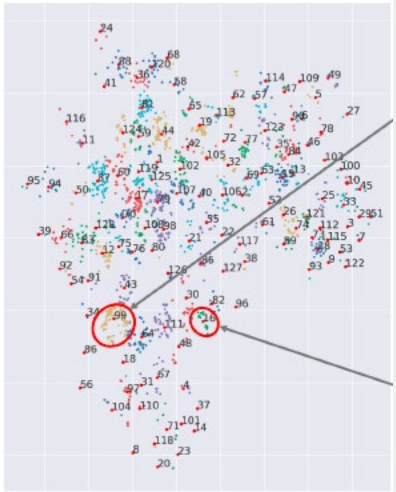


Figure 8: **EXAMPLE OF LEARNED AUDIT SAMPLES AND REPRESENTED PAYMENTS** [36]



Sample Size n = 128

Learned Representative Sample (99)

FY	Date	Payment	Department	Typ	Object	Vendor	Details	Amount USD
2017	20/04/17	CHEK17121800	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PART	G.L. SAYRE INC.	auto parts	132.59

Represented Philadelphia City Payments

FY	Date	Payment	Department	Typ	Object	Vendor	Details	Amount USD
2017	14/07/16	CHEK17005079	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	TRANSTECK INC	auto parts	–32.17
2017	20/04/17	CHEK17121800	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	G.L. SAYRE INC.	auto parts	132.59
2017	04/05/17	CHEK17125451	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	TRANSTECK INC	auto parts	37.60
2017	01/06/17	CHEK17136892	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	CHAPMAN FORD	auto parts	–293.71
2017	01/06/17	CHEK17138737	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	JACK DOHENY INC	auto parts	539.70
2017	08/06/17	CHEK17139311	25 FLEET MANAGEMENT	03 MATERIALS & SUPPLIES	MOTOR VEHICLE PARTS	TRANSTECK INC	auto parts	11.46
...	...	...	...	...	...	...	...	...

Learned Representative Sample (16)

FY	Date	Payment	Department	Typ	Object	Vendor	Details	Amount USD
2017	23/06/17	ACHD17189704	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	363.2

Represented Philadelphia City Payments

FY	Date	Payment	Department	Typ	Object	Vendor	Details	Amount USD
2017	22/07/16	ACHD17011732	13 FIRE	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	228.68
2017	23/06/17	ACHD17189704	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	113.00
2017	19/08/16	ACHD17026108	52 FREE LIBRARY	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	354.20
2017	28/10/16	ACHD17059715	22 HUMAN SERVICES	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	539.20
2017	23/12/16	ACHD17090986	42 COMMERCE	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	213.00
2017	17/02/17	ACHD17121338	22 HUMAN SERVICES	02 PURCHASE OF SERVICES	TRANSPORTATION 0211	AMERICAN EXPRESS	AMEX vouchers	320.60
...	...	...	...	...	...	...	...	...

ability to reconstruct the majority of the original city payment entries.

Figure 8 (above, left) depicts the result of a learning-based audit sample of size n = 128 of the total 238,894 city payments. Above right in *Figure 8* the learning-based audit samples #99 and #16 as well as the corresponding represented city payments are displayed. The analysis shows that *audit sample #16* represents payments by different city departments to purchase "transportation services" that were settled by "American Express Travel Vouchers" (red frames, bottom-right). *Audit sample #99* represents payments by the "Fleet Management" department, which correspond to the purchase of "vehicle parts and accessories" (red frames, top-right).

In summary, the introduced VQ-VAE based audit sampling exhibits the following advantages when compared to traditional sampling techniques (introduced in section 2):

→ *Audit Sample Representativeness:* The reconstruction error determines the audit sample's representativeness of the city payment population. The larger the distance between a given sample and its assigned payment representation, the less accurately it can be reconstructed. The level of representativeness can be adjusted based on the auditor's professional judgment. The increased sample size results in a learning-based sample representing more detailed payment characteristics and relationships (cf. *Figure 7*, sample size n = 32 vs. n = 64).

→ *Audit Sample Interpretability:* The spatial arrangements of representations allow for a human interpretation reflecting the semantic qualities of the original payment entries. Thereby, semantically similar payments reside in spatial proximity; dissimilar payments exhibit a high distance (see *Figure 8*, red circles and colour coding). Using such a holistic visual perspective provides auditors with a starting point for a targeted sampling to assess a population's homogeneity (high-density areas) and potential deviations (low-density areas).

6. CONCLUSION AND OUTLOOK

Techniques derived from DL enable auditors to draw learning-based representative samples from large-scale journal entry data. This capability is beneficial in situations where only limited or no prior information about a financial statement line item is available. Compared to traditional statistical sampling procedures, such learning mitigates the risk of selecting inadequate sampling parameters, e.g. assuming unsuitable statistical distributions or descriptive statistics. It is of particular relevance for audit practice that the representativeness of a learning-based audit sample can be determined by the reconstruction error. Similar to statistical sampling procedures, auditors can define an appropriate level of sampling risk based on their professional judgment by adjusting the sample size (see *Figure 7*). Another capability unique to DL-based audit sampling is the provision of learning-based interpretable journal entry representations (see *Figure 8*). Due to the learned dimensionality reduction, auditors are able, potentially for the first time, to visually interpret the semantics of the population to be audited. It furthermore enables auditors to address risks of material misstatements and fraud in a holistically targeted manner.

In conclusion, sample-based auditing will remain a central method for efficiently conducting audit procedures [37]. Whether the introduced DL-enabled audit sampling technique should be classified as random or systematic, or even as a separate sampling category, requires further discussion. The application of DL to obtain learning-based representative audit samples highlights the ongoing transformation of auditing from a computer-assisted practice towards an AI-augmented practice [38]. The practical implementation of DL-enabled audit sampling demonstrates the extent to which AI can contribute to the future of auditing [39]. ■

Notes: 1) International Accounting Standard (IAS) 1, Presentation of Financial Statements, International Federation of Accountants (IFAC), 2020. **2)** S. V. Grabski, S. A. Leech, and P. J. Schmidt. A review of ERP Research: A future agenda for accounting information systems. Journal of information systems, 25 (1): 37–78, 2011. **3)** Expertsuisse, Handbuch der Wirtschaftsprüfung, Band: Ordentliche Revision, 2016. **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 53rd 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 54th Hawaii International Conference on System Sciences (p. 5748), 2021. **9)** D. M. Guy, D. R. Carmichael und O. R. Whittington, Audit Sampling – An Introduction, John Wiley & Sons, 2002. **10)** See f.n. 3. **11)** See f.n. 3. **12)** G. Jokovich, Statistical Sampling in Auditing, International Journal of Accounting and Financial Management, 16: 892–898, 2013. **13)** D. M. Guy, D. R. Carmichael und O. R. Whittington, Audit Sampling – An Introduction, John Wiley & Sons, 2002. **14)** See f.n. 3. **15)** T. F. Ruud, K. Schramm und A. Allgaier, Leitlinie zum Internen Audit, Institute of Internal Auditing (IIA) Switzerland, 4. Auflage, 2021. **16)** S. Russell und P. Norvig, Artificial Intelligence: A Modern Approach, 3rd Edition, Pearson Education Ltd., pp. 1–5, 2016. **17)** D. Borth, Introduction to Machine Learning and Deep Learning, Lecture at the University of St. Gallen (HSG), 2020. **18)** 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. **19)** C. M. Bishop, Pattern Recognition and Machine Learning, Springer Science LLC, pp. 3–4, 2006. **20)** See f.n. 16. **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)** Y. LeCun, Y. Bengio, and G. Hinton, Deep Learning, Nature, 521 (7553), pp. 436–444, 2015. **23)** See f.n. 20. **24)** See f.n. 18. **25)** See f.n. 20. **26)** See f.n. 21. **27)** See f.n. 5. **28)** G. E. Hinton und R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, Science, 313(5786), pp. 504–507, 2006. **29)** 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. **30)** See f.n. 28. **31)** See f.n. 4. **32)** A. van den Oord, O. Vinyals und K. Kavukcuoglu, Neural Discrete Representation Learning, In Proceedings of the 31st 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 Quantized Variational Autoencoder Neural Networks, In Proceedings of the International Conference on Artificial Intelligence (ICAIF)'20, Association of Computing Machinery (ACM), 2020. The full study results are available through the ACM Digital Library through the following URL (viewed on 28/12/2021): https://dl.acm.org/doi/abs/10.1145/3383455.3422546. **34)** The full dataset is available through the City of Philadelphia's data portal through the following URL (viewed on 12/28/2021): https://www.phila.gov/2019-03-29-philadelphias-initial-release-of-city-payments-data/. **35)** See f.n. 33. **36)** See f.n. 33. **37)** See f.n. 3. **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.