Received: 11.10.2023 | Revised: 16.3.2024 | Accepted: 21.3.2024

A Method to Categorize and Classify Artificial Intelligence Applicable to the Risk-Based Audit Approach

Tassilo L. Föhr

Abstract: The technological capabilities of Artificial Intelligence (AI) have significant implications for the auditing profession. Nevertheless, there is no guidance on the use of AI, which is preventing its widespread adoption within the profession. To make use of AI's functionalities, there is a need for guidance on how AI can be used to gather sufficient and appropriate audit evidence. Thus, utilizing Design Science Research, the present study developed the AI Categorization and Classification (AI-CC) Method as a central artifact to provide guidance on the use of AI within the profession. The target users of the AI-CC Method are regulators, standard setters, the strategic management of the Big Four, and individual auditors. A comprehensive evaluation involving several AI experts across Germany confirmed the overall usefulness of the AI-CC Method for the entire auditing profession.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Risk-based Auditing, Audit Standards, Auditing

Eine Methode zur Kategorisierung und Klassifizierung von Künstlicher Intelligenz für den risikoorientierten Prüfungsansatz

Zusammenfassung: Die technologischen Möglichkeiten der Künstlichen Intelligenz (KI) haben erhebliche Auswirkungen auf den Berufsstand der Wirtschaftsprüfer. Dennoch gibt es keine Leitlinien für den Einsatz von KI. Dies verhindert ihre breite Anwendung in der Branche. Um die Funktionalitäten der KI nutzen zu können, bedarf es einer Anleitung, wie KI eingesetzt werden kann, um ausreichende und angemessene Prüfungsnachweise zu sammeln. Daher wurde in der vorliegenden Studie unter Verwendung von Design Science Research die KI-Kategorisierungs- und Klassifizierungsmethode als zentrales Artefakt entwickelt, um eine Anleitung für den Einsatz von KI in der Praxis zu geben. Zielgruppen der Methode sind Regulierungsbehörden, Standardsetzer, das strategische Management der Big Four und einzelne Prüfer. Eine umfassende Evaluierung, an der mehrere KI-Experten aus ganz Deutschland beteiligt waren, bestätigte den allgemeinen Nutzen der Methode für den gesamten Berufsstand der Wirtschaftsprüfer.

Stichworte: Künstliche Intelligenz, Maschinelles Lernen, Tiefes Lernen, risikoorientierter Prüfungsansatz, Prüfungsstandards, Wirtschaftsprüfung

1 Introduction

The overarching goal of a financial statement audit is to enable the auditor to derive an audit judgment concerning the degree to which the financial statements are aligned with pertinent accounting standards (International Standard on Auditing (ISA) 200.5 and 200.11 (a), IFAC, 2009a). Thus, the initial performance metric of concern in the auditing process is its effectiveness, which is defined of issuing an audit judgment with reasonable assurance, and another performance metric of concern is the efficiency of the audit procedures to be conducted, meaning that the most cost-effective auditing procedures should be selected and implemented (Marten et al., 2020). Consequently, the primary objective is issuing an audit judgment with reasonable assurance while minimizing the audit costs, in compliance with the existing ISA 200 (IFAC, 2009a). It has been shown that this primary objective of a risk-based audit can be achieved by utilizing Artificial Intelligence (AI) (Fedyk et al., 2022).

In today's technologically evolving environments, auditors' tools and techniques have been subjected to significant changes based on emerging relevant audit information from internal and external sources (Vasarhelyi et al., 2015; Warren et al., 2015). Furthermore, increasing processing power of cloud systems, graphics processing units (GPUs), and evolving software resources, yield to a rise of AI (Kokina & Davenport, 2017) which has significantly impacted the auditing profession (Austin et al., 2021; Feliciano & Quick, 2022). Studies suggest that by 2026, organizations will experience an improvement of over 50 % in operationalizing AI (Gartner, 2024) and by 2025 approximately 30 % of corporate audits will utilize AI-based judgement support (World Economic Forum, 2015). Thus, auditing firms are already increasingly focusing on the development of AI-augmented auditing techniques (World Economic Forum, 2015). It seems challenging for an auditor to gather sufficient appropriate audit evidence by relying merely on traditional approaches, without utilizing AI (Kokina & Davenport, 2017). More than ever, auditors are challenged to advance their technological capabilities (Warren et al., 2015). Therefore, auditors shall reconsider existing auditing procedures or combine them with AI-Technologies[1] to meet the key performance metrics of effectiveness and efficiency in issuing audit judgments with reasonable assurance (Issa et al., 2016).

Nevertheless, there are currently major barriers to the widespread adoption of AI within the auditing profession. Although some researchers have identified missing guidance for auditors in this regard as one of such barriers (Austin et al., 2021; Christ et al., 2021), there has been little response to this matter to date. Recent studies on AI have focused on highly specialized use cases, such as fraud detection (Schreyer et al., 2018), contradiction detection (Deußer et al., 2023), KPI extraction (Hillebrand et al., 2022), and audit sampling (Schreyer et al., 2020), and there is still no systematic guidance in identifying the most promising AI-Technologies for gathering sufficient and appropriate audit evidence. As a result, auditing firms are endeavoring to utilize AI in real-world audit settings based only on their own empirical experiences (Brenner et al., 2020; Salijeni et al., 2021), and the application of AI in practice remains in the research and development stage (Gierbl et al., 2021; Krieger et al., 2021).

1 The term "AI-Technologies" is used to address several methods and algorithms, including AI, Machine Learning, or Deep Learning. The paper always refers to these technologies using the term "AI-Technologies".

As a first step to overcoming the barriers to AI implementation, the International Auditing and Assurance Standards Board (IAASB) established a technology consultation group that provides auditors with non-authoritative guidance in utilizing AI in addition to the existing ISAs (IAASB, 2024). Furthermore, the IAASB clarified that the existing risk-based audit approach does not need to be fundamentally modernized because the underlying ISAs are principle-based and therefore allow the use of AI (IAASB, 2022). The present study aimed to build on these recent developments in providing the auditing profession with additional guidance in utilizing AI.

Due to the complexity of AI-Technologies and the variability of several risk-based audit procedures that are necessary during an audit, the assessment of AI-Technologies to determine the most promising among them for gathering sufficient and appropriate audit evidence is not straightforward. Furthermore, existing research and modernized audit standards, such as ISA 315 (Revised 2019) (IAASB, 2019) do not give concrete assessment guidance to practitioners. Thus, there has been limited systematic assessment of the applicability of AI-Technologies to the risk-based audit approach. The present study sought to solve this real-world problem by answering the following research question: *How can the applicability of representative AI-Technologies to the existing risk-based audit approach be systematically assessed?*

To solve the aforementioned real-world problem, the present study proposes the AI Categorization and Classification (AI-CC) Method as its central artifact, based on the design science research (DSR) paradigm (Hevner et al., 2004; Gregor & Hevner, 2013). The AI-CC Method consists of two components that technically define the artifact as a method: (1) a categorization framework with dimensions, requirements, and analytics as components and (2) the assessment process describing how to utilize the artifact. The method was designed to enable organizations' units responsible for implementing AI (Fedyk et al., 2022), such as regulators, the strategic management of the Big Four, and individual auditors from smaller auditing firms, to systematically assess AI-Technologies' capabilities to the risk-based audit approach. Thus, the developed artifact increases the chances of attaining the primary objective of an audit by identifying AI-Technologies that can best foster effectiveness and efficiency in audit procedures. It allows for a more objective assessment of the best AI-Technologies to use for auditing purposes. Furthermore, the developed AI-CC Method illustrates a research-based solution with practical implications for determining when and when not to utilize a specific AI-Technology to gather audit evidence. Therefore, the method addresses the explicitly stated problem of lack of guidance in utilizing AI within the auditing profession and thus responds to the calls for more practically relevant research contributions (Bennis & O'Toole, 2005; Corley & Gioia, 2011; McCarthy, 2012; Waymire, 2012; Wood, 2016; Summers & Wood, 2017; Peffers et al., 2018; Moon & Wood, 2020; Burton et al., 2021; Rajgopal, 2021; Burton, Heninger, et al., 2022; Burton, Summers, et al., 2022).

This paper was structured according to the DSR methodology of Peffers et al. (2007). Having identified the real-world problem and defined the overall objectives of the AI-CC Method, the paper provides a structured literature review in section 2 and describes the research design for the development of the method in section 3. Section 4 explains the method, section 5 demonstrates its use, and section 6 presents a detailed evaluation of it. Section 7 discusses the study's contributions, and section 8 closes the paper with a conclusion.

2 Literature Review

Through a structured literature review, the present study ensured the novelty of the developed artifact by confirming that the AI-CC Method or a similar method had not been previously developed under the lens of DSR. Based on the requirements for the rigor of a literature search, the structured literature review (see Appendix I) identified previous studies that utilized DSR in the field of emerging technologies in auditing and accounting (vom Brocke et al., 2009; vom Brocke et al., 2015). The aim of the review was to identify DSR research contributions regarding advanced Data Analytics employing AI, Machine Learning (ML), or Deep Learning (DL), and additional emerging technologies affecting the auditing and accounting profession. This more general view on emerging technologies was adopted because of the novelty and rarity of research contributions in the field of auditing and accounting that have been explicitly developed under the DSR paradigm. Furthermore, the scope of the search was broadened to identify existing artifacts with functions similar to those of the AI-CC Method. The identified DSR contributions were supplemented by implicit DSR contributions from the existing literature to ensure the selection of studies on artifacts representing the current state-of-the-art.

2.1 Explicit DSR Contributions

There is a major body of DSR contributions focusing on the enhancement of risk-based audits of financial statements with emerging technologies. One focus area of audit phases enhanced, especially by AI and advanced Data Analytics is *initial risk assessment*. Some studies have focused on information integration by combining XBRL data with textual data extracted from MD&A to foster decision-making (Chou et al., 2016), advanced expert systems (Lombardi & Dull, 2016), and enhanced fraud risk assessment by utilizing geographical data in the form of spherical distances (Huang et al., 2022). Other studies have focused on cluster analysis of governmental data (Alzamil et al., 2021) or analyzing the textual contents of annual reports to predict financial performance (Mousa et al., 2022). In addition, Robotic Process Automation (RPA) has been shown to be capable of automating repetitive tasks, enabling auditors to devote more time to carrying out risk assessment procedures (C. A. Zhang et al., 2022). Eulerich et al. (2022) provided detailed guidance on how to identify the most suitable tasks to automate with RPA.

Observing inventory counts through drone-enabled technology (Appelbaum & Nehmer, 2017), improving in the detection of material weaknesses in internal controls by applying ML (Nasir et al., 2021), and process mining studies utilizing journal entry data, taking into account the booking logic in Enterprise Resource Planning (ERP) Systems to improve internal control evaluation (Werner & Gehrke, 2015; Werner, 2017, Werner, 2019; Werner & Gehrke, 2019), are recent DSR contributions attributable to *internal control evaluation*.

Substantive Procedures are also covered by recent DSR contributions. Frameworks have been developed and used to identify and investigate severe outliers in financial data (No et al., 2019; Freiman et al., 2022) or to improve audit quality by utilizing drone-enabled technology with subsequent object detection models (Christ et al., 2021). Furthermore, visualization techniques in the form of accounting graph typologies, which can enhance fraud detection, have been developed (Guo et al., 2022) along with methods based on belief functions that are capable of evaluating audit evidence for control testing (Nehmer & Srivastava, 2016).

Beyond the phases of a risk-based audit, there have also been data-based DSR contributions focusing on consistency checks among qualitative and quantitative data (Chou et al., 2018) or audit procedures over XBRL-tagged datasets (Boritz & No, 2016). Some further DSR contributions in the field of emerging technologies focused on methods for Continuous Auditing (Kiesow et al., 2015; Polizzi & Scannella, 2023) and frameworks with subsequent anomaly detection methods to mitigate information overload (Perols & Murthy, 2012; G. Zhang et al., 2022). Other DSR contributions focused on IT security or privacy issues resulting from the use of cloud-based systems (Singh & Dutta, 2018; Coss & Dhillon, 2020) and from the entire IT landscape of an organization (Otero, 2015; Rahimian et al., 2016; Brunner et al., 2018; Kogan & Yin, 2021; Plant et al., 2022). The emerging field of Blockchain technologies has also been researched under the prism of DSR, with a focus on building and implementing of such technologies (O'Leary, 2019; Rozario & Thomas, 2019; Centobelli et al., 2022) and on the integrity and reliability issues resulting from the utilization of Blockchains (McCallig et al., 2019; Appelbaum & Nehmer, 2020; Albizri & Appelbaum, 2021; Sheldon, 2021; Ritchi et al., 2022). Furthermore, M. Liu et al. (2021) investigated the impact of Blockchains on the auditing and accounting profession.

2.2 Implicit DSR Contributions

There are some implicit DSR anchors, especially in the field of the use of AI in risk-based auditing. The following steps can be taken to abstractly determine the fulfillment of DSR requirements: (1) identifying a significant audit problem; (2) designing and developing an innovative artifact to solve the problem; and (3) validating that the developed artifact is capable of solving the auditing problem (Kogan et al., 2019). Some implicit DSR contributions are presented in this section to further confirm the novelty of the AI-CC Method.

Regarding unstructured data sources, AI can enhance *risk assessment* by extracting valuable content from audit brainstorming sessions as decision support for auditors (Li & Liu, 2020). Furthermore, AI is capable of extracting sentiments from annual reports (Azimi & Agrawal, 2021), key audit matters (Liu et al., 2022), and quarterly earnings disclosures (Siano & Wysocki, 2021) to identify risky areas. Special AI-Technologies have also been developed for financial text data, with functionalities outperforming those of more traditional approaches (Liu et al., 2020; Huang et al., 2023). *Risk assessment* can also be supplemented by AI analysis of structured data to detect financial misstatements in transactions (Bertomeu et al., 2021), by applying clustering mechanisms (Byrnes, 2019), or by providing decision support for auditors focused on fraud risks (Hooda et al., 2020). *Internal control evaluation* can also be enhanced by AI-based decision support systems for the detection of material weaknesses in internal controls (Sun, 2019; Zhaokai & Moffitt, 2019). Furthermore, analyzing conference calls by extracting sentiments using AI can identify material control weaknesses (Sun, 2018). Existing research also shows how AI can supplement *substantive procedures* by enhancing fraud detection (Schreyer et al., 2018; Bao et al., 2020) or outlier identification (No et al., 2019) among structured financial data. During *audit completion*, AI can foster audit opinion prediction based on the clients' financial statements (Saeedi, 2021), and can supplement the audit of disclosures in the notes by matching text passages with specific laws or regulations (Sifa et al., 2019).

3 Research Design

The present study adopted the DSR paradigm (Hevner et al., 2004; Gregor & Hevner, 2013) and followed the DSR methodology (Peffers et al., 2007) to develop the AI-CC Method as its central artifact. An overview of the study's research design is provided in Figure 1.

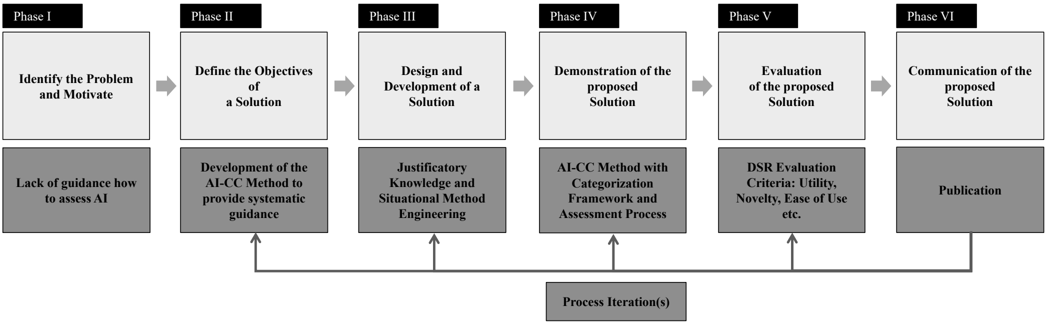

Figure 1: DSR Methodology based on Peffers et al. (2007)

The first two research phases have already been illustrated in the Introduction section. The specifications of the developed artifact were decided during the design and development phase. The AI-CC Method, with its underlying categorization framework and assessment process, can be technically defined as a structured and systematic method for solving existing real-world problems (Avison, 1996; Brinkkemper, 1996). Thus, it had to be developed based on justificatory knowledge (Gregor & Jones, 2007) relevant for the development of a method to ensure its rigor (Hevner et al., 2004). In the study's setting, the components of the AI-CC Method were defined based on the study of Braun et al. (2005), which provided systematic procedures that could be used to achieve the defined objectives of the artifact. The method was further detailed by its underlying attributes (goal, systematic structure, principles, and repeatability mechanisms) and elements (metamodel, role, technique, activity, tool, and output). Detailed information about the specific method components is shown in Table 1.

Situational Method Engineering (SME) was used as a research method to design and construct the AI-CC Method and to adapt it to or make it suitable for specific situations (Brinkkemper, 1996; Henderson-Sellers & Ralyté, 2010). The assembly-based approach from Ralyté et al. (2003) was used for the development of the method; parts of existing methods that are necessary to compose the final artifact were compiled. In line with the overall objectives and use cases addressed by the method, the final artifact consists of a categorization framework with dimensions, requirements, and analytics components representing a metamodel and an assessment process for the artifact's execution. The categorization framework's dimensions were defined using the Risk-based Audit Approach Classification System based on existing ISAs and the AI-Technology Classification System derived from AI sourcebooks. Furthermore, the requirements component was defined

	Component	Detailed Information
Attributes	**Goal**	The method forms the basis for achieving goals.
	Systematic structure	The method enables concrete work procedures and tasks for achieving goals.
	Principles	The method is based on construction guidelines and strategies.
	Repeatability	The method is repeatable in different settings.
Elements	**Metamodel**	The method includes a metamodel for conceptualizing the results.
	Role	The method incorporates activities performed based on different roles (e.g., people, hierarchical position).
	Technique	The method provides instructions on how to perform specific activities.
	Activity	The method incorporates construction tasks for achieving specific results.
	Tool	The method utilizes tools in support of the implementation and application of techniques.
	Output	The method has outputs that define its results.

Table 1: Components of the AI-CC Method based on Braun et al. (2005), Vanwersch et al. (2016), and Denner et al. (2018)

using functional requirements based on audit assertions (ISA 315.A190 (Revised 2019), IAASB, 2019) and nonfunctional requirements based on the quality-in-use model and the product quality model illustrated in the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 25010 (ISO/IEC, 2011, 2023a)[2]. The analytics component was compiled by implementing the weighted degree of context

2 The ISO/IEC 25010:2011, (ISO/IEC, 2011) has been revised after the development and evaluation of the AI-CC Method by ISO/IEC 25002:2024, (ISO/IEC, 2024), ISO/IEC 25010:2023, (ISO/IEC, 2023a), and ISO/IEC 25019:2023, (ISO/IEC, 2023b). The AI-CC Method is initially based on the requirements of the Quality-in-use model and Product quality model detailed in ISO 25010:2011, (ISO/IEC, 2011). However, the basic purpose of these two models has not been changed due to the revision. Instead, the revision's primary objective was to update and refine terminology for a better understanding of these two models.

specificity (DCS) and the total value of applicability (TVA). The DCS indicator initially developed by vom Brocke et al. (2021) was adopted to meet the specified objectives of the AI-CC Method (Gregor & Hevner, 2013). The concrete execution order of the AI-CC Method is defined by the assessment process. The assessment process executes the existing method fragments of the categorization framework based on the artifact's overall objectives, relying on multiple classification techniques. The logic of the assessment process was initially based on the Context-Aware Business Process Management Method (BPM) Assessment and Selection (CAMAS) Method (vom Brocke et al., 2021) for context-aware BPM as justificatory knowledge. However, the AI-CC Method was furnished with several additional components to ensure its innovativeness.

To demonstrate the AI-CC Method we explain its design specifications in detail in the next section. A Microsoft Excel (MS Excel) prototype of the AI-CC Method with a representative sample of AI-Technologies[3] supporting the execution of the assessment process is also provided as an electronic supplementary material.

To evaluate the AI-CC Method, evaluation mechanisms focusing on established DSR evaluation criteria (e.g., utility, novelty and importance, ease of use, and performance) were used (March & Smith, 1995; Hevner et al., 2004; Sonnenberg & vom Brocke, 2012). The artifact was evaluated by two internal researchers and 11 AI experts across Germany who executed the assessment process in line with the artifact's categorization framework.

4 Artifact Description

4.1 AI-CC Method Overview

The AI-CC Method consists of a categorization framework and an assessment process. Mapping the method based on its components in Table 1 illustrates the rigor of the method's design specifications, focusing on the systematic assessment of AI-Technologies in terms of their applicability to the risk-based audit approach. This systematic assessment identifies the most promising AI-Technologies for obtaining sufficient and appropriate audit evidence (*goal*). The categorization framework structures the assessment of AI-Technologies based on its dimensions, requirements, and analytics components (*principles*) specified during the development stage. Guidelines on how to execute the categorization framework are embedded in the assessment process (*systematic structure*). Successfully assessed AI-Technologies are integrated into the method's AI-Technology base.

The assessment process of the AI-CC Method can be further specified by techniques, tools, roles, and outputs, which ensure that the method is executable and applicable in various circumstances across different end users (*repeatability*). Figure 2 provides a comprehensive overview of the method and the interdependencies between the categorization framework and the assessment process. Detailed information about the method is presented in the following chapters. To better understand the working mechanisms of the developed artifact, please find the blank version of the AI-CC Method's MS Excel prototype as an electronic supplementary material.

3 To receive a copy of the MS Excel prototype of the AI-CC Method, please contact the author or the Institute of Accounting and Auditing at the University of Ulm.

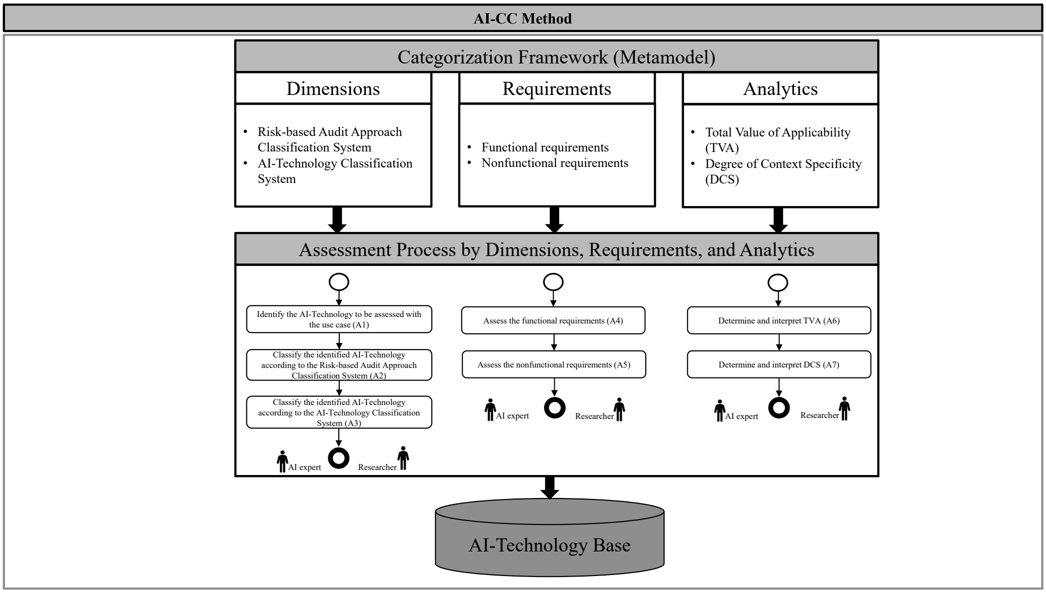

Figure 2: AI-CC Method based on vom Brocke et al. (2021)

4.2 Categorization Framework

The core element of the AI-CC Method is its categorization framework, illustrated in Figure 3, which serves as a metamodel for the execution of the assessment process. The categorization framework helps in the systematic assessment of AI-Technologies based on certain dimensions, requirements, and analytics.

Dimensions are defined as multi-criteria classification mechanisms for assessing AI-Technologies along two components originating from auditing and technical perspectives. Dimensions needed to be set to ensure the existence of a solid base of systematically classified AI-Technologies before the assessment of the requirements. The AI-CC Method can be used to systematically assess AI-Technologies, because it covers both categorization and classification by dimension, requirement, and analytics.

The Risk-based Audit Approach Classification System defines the audit perspective as stated in the existing risk-based audit approach and therefore allows for the classification of AI-Technologies by audit procedure, regulatory and legal audit framework, and audit phase. The audit procedure elements of the categorization framework focus on classification according to the concrete audit procedures addressed by AI-Technologies, such as risk assessment and control evaluation (ISA 315 (Revised 2019), IAASB, 2019), substantive analytical procedures (ISA 520, IFAC, 2009d), and fraud detection (ISA 240, IFAC, 2009b). The regulatory and legal audit framework refers to the ISA standards relevant to each audit procedure addressed by an AI-Technology. Audit phases focus on the phases of a risk-based audit, and an AI-Technology can be classified as one for audit planning and risk assessment, substantive procedures, audit review, or audit completion.

The AI-Technology Classification System represents the technical perspective of the categorization framework. Therefore, AI-Technologies can be classified by data structure, domain, classification, paradigm, method, and approach. The data structure can be defined as the relevant data modalities during a risk-based audit in structured and unstructured forms. The domain level defines the procedures necessary to implement the functionalities of AI in conducting risk-based audit procedures, such as Data Mining, Text Mining, Natural Language Processing, Computer Vision, and Process Mining[4]. The classification component distinguishes AI-Technologies in a more detailed way: AI in general, traditional ML, and DL. The paradigm component refers to the specific learning procedures of supervised and unsupervised paradigms. The method component focuses more on the problem types addressed by AI-Technologies, such as classification, regression, clustering, outlier detection, decision support, process visualization, prediction, and object detection. Finally, the approach component details the approaches that enable AI-Technologies to solve problems based on the previously determined technical components.[5]

The requirements component focuses on the given nature of an AI-Technology and cannot be modified. The component consists of functional and nonfunctional requirements and therefore represents concrete factors for assessing AI-Technologies, referring to audit-related requirements (functional) and technical-related requirements (nonfunctional). To provide unified measures and definitions for these requirements, the functional part was based on audit assertions, such as occurrence, accuracy, cutoff, and existence (ISA 315.A190 (Revised 2019), IAASB, 2019). It is the systematic assessment of these audit assertions that enables the identification of the most promising AI-Technologies for obtaining sufficient and appropriate audit evidence. The nonfunctional requirements are defined by ISO/IEC 25010 (ISO/IEC, 2011, 2023a) and ISO/IEC 25019 (ISO/IEC, 2023b) covering essential factors for assessing the functionality of AI-Technologies, such as effectiveness, efficiency, functional suitability, and usability.[6] In contrast to the method developed by vom Brocke et al. (2021), the AI-CC Method allows for a more nuanced rating on an ordinal scale so that end users can rate more distinguishable factors related to each requirement. The categorization framework defines this assessment scheme as the degree of applicability.

The analytics component provides valuable insights from the assessment of AI-Technologies. The main elements of this component are TVA and weighted DCS. TVA helps in identifying promising AI-Technologies for the auditing profession, because a high TVA indicates a high degree of applicability. DCS, on the other hand, provides insights into the generality or specificity of an AI-Technology according to the stated requirements. The next section provides more details about the analytics component.

4 The focus of the AI-CC Method is the systematic assessment of AI-Technologies. Nevertheless, it was shown in the study that the AI-CC Method is also applicable to emerging technologies as the study included Process Mining as a representative example applicable to systematic assessment by the AI-CC Method during the evaluation of the method.

5 More technical details about the AI-Technology Classification System are provided in Appendix II.

6 More technical details about the functional and nonfunctional requirements are provided in Appendix III.

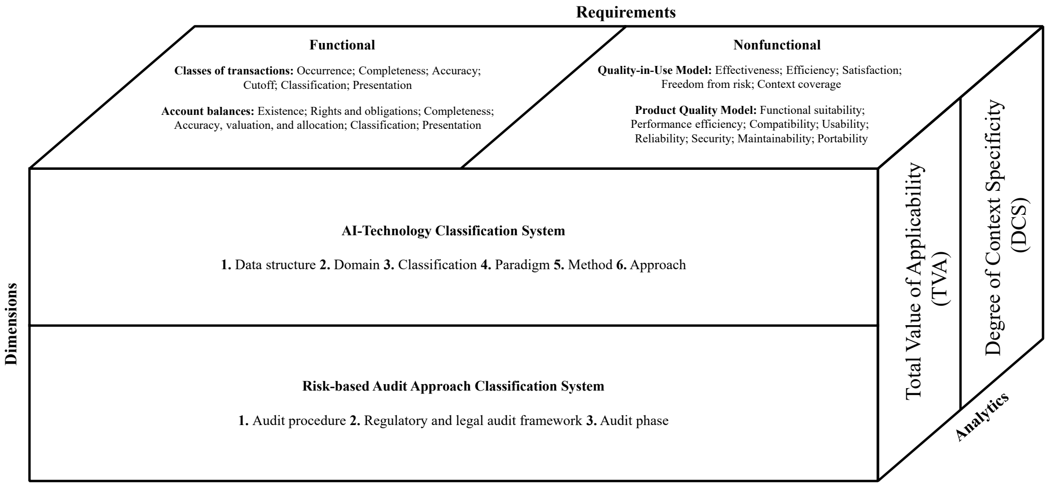

Figure 3: Categorization Framework (Metamodel) based on vom Brocke et al. (2021)

4.3 Assessment Process

The assessment process ensures the existence of relevant method attributes (Table 1) because it provides end users with detailed guidance on how to assess AI-Technologies (*goal*) based on the components of the categorization framework (*principles*). The assessment process consists of seven consecutive steps aggregated and combined in terms of dimensions, requirements, and analytics (*systematic structure*), as shown in Table 2. Completing each step ensures the correct execution of the AI-CC Method during different situations across various end users (*repeatability*).

Identifying an AI-Technology to assess in line with the categorization framework (Activity 1) is the first process step (*technique/output*). A systematic literature search (see Appendix IV) can help in identifying the AI-Technology suitable for assessment (*tool*). To ensure that the specific AI-Technology can be systematically assessed (*technique*), the theoretical background and definition of AI-Technologies, and the link to risk-based audit procedures must be used as bases (*tool*). This process step can be executed by interested researchers or several stakeholders, such as standard setters, auditors, or clients who want to identify the most promising AI-Technologies for a risk-based audit (*role*).

Classifying the identified AI-Technology according to the Risk-based Audit Approach Classification System (Activity 2) ensures a link to the risk-based audit approach, as defined in the categorization framework (*technique/output*). The same AI-Technologies can be assessed to different alternatives of the Risk-based Audit Approach Classification System as illustrated in the working mechanisms of the MS Excel prototype (*tool*). This activity can be performed by researchers or the previously defined stakeholders (*role*).

Classifying the identified AI-Technology according to the AI-Technology Classification System (Activity 3) involves classifying the AI-Technology according to its technical characteristics (*technique/output*) defined in the categorization framework. This activity can be documented within the MS Excel prototype of the developed artifact (*tool*) and can be executed by the previously defined end users (*role*).

Activity	Technique	Tool	Output	Role
Identify the AI-Technology to be assessed with the use case (A1).	Identify an AI-Technology for a risk-based audit procedure from the existing literature.	Literature review, Definition of AI	AI-Technology with use case as a basis	AI expert or researcher
Classify the identified AI-Technology according to the Risk-based Audit Approach Classification System (A2).	Classify the AI-Technology in terms of audit procedure, regulatory and legal audit framework, and audit phase.	Categorization Framework, Risk-based Audit Approach Classification System	AI-Technology classified in terms of the Risk-based Audit Approach Classification System	AI expert or researcher
Classify the identified AI-Technology according to the AI-Technology Classification System (A3).	Classify the AI-Technology according to the data structure, domain, classification, paradigm, method, and approach.	Categorization Framework, AI-Technology Classification System	AI-Technology classified in terms of the AI-Technology Classification System	AI expert or researcher
Assess the functional and nonfunctional requirements (A4 and A5).	Classify the AI-Technology according to its functional and nonfunctional requirements.	Categorization Framework, Degree of Applicability (ordinal scale)	AI-Technology assessed based on the functional and nonfunctional requirements	AI expert or researcher
Determine and interpret the Analytics Components (A6 and A7).	Measure and assess the TVA and DCS.	Categorization Framework, Calculated values of TVA and DCS	AI-Technology assessed based on TVA, and DCS	AI expert or researcher

Table 2: Overview of Assessment Process based on vom Brocke et al. (2021)

The main part of the AI-CC Method is the assessment of AI-Technologies based on the functional and nonfunctional requirements of the categorization framework (Activities 4 and 5) (*output*). This systematic assessment indicates whether an AI-Technology meets the requirements, as measured by the so-called degree of applicability (*technique*).

The degree of applicability shall be assessed based on the following ordinal scale (*tool*):

- *0, 1, 2, 3, 4, 5*: The extent to which the AI-Technology matches the specific requirement, where 0 indicates the lowest degree of requirement fulfillment (no applicability) and 5 indicates the highest degree of requirement fulfillment (very high applicability). The proposed MS Excel prototype provides end users with the opportunity to rank an AI-Technology based on assessment criteria between 0 and 5.
- *Not assessable (–)*: The AI-Technology is not assessable in terms of specific requirements.[7]

7 The option "not assessable" is outside the ordinal scale and should be used only if there is no information for assessing the degree of applicability for a specific requirement.

Determining and interpreting the analytics component according to the categorization framework as the last process step (Activities 6 and 7) provides a detailed view of the assessed AI-Technology (*output*). TVA aggregates the values of all the assessed requirements and thus identifies the AI-Technology with the highest functionality (*technique/tool*). The weighted DCS indicates whether the AI-Technology assessed has a more general or specific functionality. This ensures the identification of AI-Technologies that can effectively and efficiently serve specific needs in the auditing field. The formula for weighted DCS is based on vom Brocke et al. (2021) and shown in equation (1) below:

$$DCS = \left(1 - \left(\frac{\sum_{f \in F} \frac{\sum_{f \in C_f} \delta_f}{|C_f|}}{(|F| * 5)}\right)\right) * \left(1 - \frac{\beta}{\sum_{f \in F} |C_f|}\right) \quad (1),$$

where F is the set of functional and nonfunctional requirements included in the artifact, C_f is the set of characteristics per requirement set $f \in F$, δ_f is the set of requirements assessed with 0, 1, 2, 3, 4, or 5 for each requirement set $f \in F$, and β is the number of requirements determined as "not assessable" (–) across all characteristics C_f of the set of functional and nonfunctional requirements $f \in F$. A DCS of approximately 100 % means that the AI-Technology meets only one requirement per requirement set, whereas a DCS of approximately 0 % means that the AI-Technology is a more general-purpose technology (*technique/tool*). This process step can be performed by researchers and relevant stakeholders (*role*).

5 Demonstration

To demonstrate how to implement the AI-CC Method illustrated in Figure 2, and to show the artifact's functionalities, the method's assessment process (Table 2) in line with its categorization framework was applied to a representative example. The concrete demonstration was conducted by two researchers[8] specializing in the field of AI in auditing and accounting. They discussed their views and opinions regarding the systematic assessment by dimension, requirement, and analytics. They agreed on only one representative assessment of the chosen AI-Technology. The combined representative assessment results are shown along with the underlying reasoning process.

5.1 Dimensions

The two researchers decided to use the AI-Technology proposed by Schreyer et al. (2018) as a representative example (Activity 1) to demonstrate the assessment process of the AI-CC Method. The proposed AI-Technology is defined as deep autoencoder networks for detecting anomalies in journal entry data, so that sufficient appropriate audit evidence can be gathered. Deep autoencoders gained attention in auditing for their capabilities to reproduce journal entries and flag anomalous data patterns (Schultz & Tropmann-Frick, 2020; Nonnenmacher et al., 2021). Both researchers decided to assess the technology according to the dimensions of the artifact's categorization framework, as presented in the next paragraph:

8 Researchers within the university, where the study took place.

Demonstrating Activity 2, Schreyer et al. (2018) showed how the audit phase of substantive procedures (ISA 330, IFAC, 2009c) can be enhanced with substantial analytical procedures (ISA 330, IFAC, 2009c; ISA 520, IFAC, 2009d), proposing the application of deep autoencoder networks focused on fraud detection in large-scale accounting data (ISA 240, IFAC, 2009b).

Activity 3 ensures the classification of the identified AI-Technology from a technical perspective according to the AI-Technology Classification System by classifying the data described as SAP ERP journal entries as structured data. Anomaly detection procedures based on journal entry data are in the domain of data mining. Deep autoencoder networks are represented by several hidden layers within their architectures, which point to the main features of DL. They are also known for their unsupervised data compression capabilities for detecting anomalies (Hawkins et al., 2002; Williams et al., 2002). Anomaly scores refer to anomaly detection methods defined by specific threshold parameters. Deep autoencoder networks are often characterized as replicator neural networks with encoder and decoder architectures (Rumelhart et al., 1987) to replicate their input data. The disparity between the original input and its replicated output is termed as the "reconstruction error" (Schreyer et al., 2018).

5.2 Requirements

To demonstrate Activity 4 and 5 as the processing and coding of the requirements in line with the categorization framework of the AI-CC Method, the combined results of the two researchers' reasoning processes are presented to determine whether the AI-Technology is applicable (value >= 3) or not applicable (value < 3) according to each requirement (see Appendix III for a detailed description of each requirement). The researchers discussed the functionality and agreed on one final value for each requirement (see Tables 3–6 for the detailed values).

As shown by Schreyer et al. (2018) (see Table 3), deep autoencoders are capable of determining whether transactions have been recorded, measured, and adjusted appropriately (*accuracy, valuation, and allocation*). "Non-anomalous" transactions can be reconstructed, whereas "anomalous" transactions result in reconstruction errors. If a transaction is posted at an unusual time, deep autoencoders will also result in reconstruction errors for these specific columns within the journal entry data (*cutoff*). As deep autoencoders also identify unusual account combinations, known as local anomalies, they can ensure the audit assertion of *classification*. Furthermore, deep autoencoders have implications for fraud detection by highlighting anomalous journal entries (*occurrence)* requiring manual investigation by an auditor.

Deep Autoencoder Applicable Functional Requirements (Rating Values)	
Occurrence	4
Accuracy	4
Cutoff	4
Classification	5
Accuracy, valuation, allocation	4

Table 3: Applicable Functional Requirements

As illustrated in Table 4, the *existence* and *completeness* requirements seem hard for the proposed deep autoencoders to grasp as there is a need for additional audit procedures ensuring these assertions. For example, to ensure the *existence* and *completeness* of inventories, some drone observation procedures could be carried out to gain sufficient and appropriate audit evidence (Christ et al., 2021) in addition to deep autoencoders analyzing journal entry data. To ensure *presentation*, the deep autoencoders need to analyze the data presented in the final report, not just journal entry data, as illustrated by Schreyer et al. (2018). The upholding of *rights and obligations* can be ensured only if further textual analysis of specific contracts (Zhaokai & Moffitt, 2019) is conducted in addition to analysis of journal entry data with deep autoencoders.

Deep Autoencoder Nonapplicable Functional Requirements (Rating Values)	
Existence	2
Completeness	2
Rights and obligations	2
Presentation	2

Table 4: Nonapplicable Functional Requirements

Regarding the assessment of an AI-Technology in terms of the nonfunctional requirements (Activity 5) (see Table 5 and 6), Schreyer et al. (2018) showed the *effectiveness, reliability*, and *functional suitability* of deep autoencoders based on the illustrated evaluation metrics (sensitivity, precision, F1 Score, top-k precision, number of detected anomalies) and low false positive rates, which are also relevant for real-world auditors because each potential anomaly requires additional investigation efforts. Deep autoencoders also obtained high assessment scores for *satisfaction* because their benchmark evaluations with traditional ML approaches showed their superior performance. The underlying model architectures have many hyperparameters, such as the number of hidden layers, which can be corrected, improved, or adapted if the current output is not the desired one (*maintainability*). The most complex version of the proposed deep autoencoders from Schreyer et al. (2018) showed the highest performance. *Compatibility* is also high because deep autoencoders function on general system components accessible via cloud services or local GPUs. The suggested graphical output illustrations of the deep autoencoders' results

Deep Autoencoder Applicable Nonfunctional Requirements (Rating Values)	
Effectiveness	4
Reliability	4
Functional suitability	4
Satisfaction	4
Maintainability	5
Compatibility	4
Usability	4

Table 5: Applicable Nonfunctional Requirements

as suggested by Schreyer et al. (2018) can also be understood by nontechnical auditors (*usability*).

Deep autoencoders have lower *efficiency* compared to traditional ML-models because of the various technical resources necessary to train them to achieve superior functionality compared to traditional ML approaches. For example, in the training stage of DL algorithms, assuming a scalable deployment setting over several audit engagements, GPUs would be necessary. This requirement is closely intertwined with the assessment of *performance efficiency* because deep autoencoders' resource consumption to achieve superior functionality is relatively high compared to traditional ML. The proposed deep autoencoders' *freedom from risk* and *security* cannot be guaranteed or assessed because all the deep autoencoders utilized by Schreyer et al. (2018) are in the prototype stage before deployment. Furthermore, it is uncertain how deep autoencoders can be adapted to different contexts due to their task-specific character, which yielded lower assessments of *context coverage* and *portability*.

Deep Autoencoder Nonapplicable Nonfunctional Requirements (Rating Values)	
Efficiency	2
Performance efficiency	2
Freedom from risk	2
Security	0
Context coverage	2
Portability	2

Table 6: Nonapplicable Nonfunctional Requirements

5.3 Analytics

In determining and interpreting the analytics components of the AI-CC Method based on the combined assessments of the two researchers (Activities 6 and 7), the deep autoencoders of Schreyer et al. (2018) yielded the analytical values illustrated in Table 7. Compared with the other AI-Technologies identified and assessed, as discussed in the following section, deep autoencoders achieved a high TVA of 77. This high value shows that deep autoencoders are suitable for various requirements. This is in accordance with the advanced functionality of DL models compared to several traditional ML models evaluated in the next section. Furthermore, the proposed deep autoencoders obtained a DCS value of 39 %, which means that, although they are task-specific models, they can address several requirements by analyzing journal entry data for anomalous transactions.

Deep Autoencoder Analytical Results	
Total Value of Applicability (TVA)	77
Degree of Context Specificity (DCS)	39 %

Table 7: TVA and DCS of Deep Autoencoders

6 Evaluation

6.1 Procedure

This section discusses the overall usefulness, effectiveness, and efficiency of the AI-CC Method in providing guidance on how representative AI-Technologies can be systematically assessed vis-à-vis the existing risk-based audit approach. To comprehensively evaluate the method, it was evaluated based on established DSR evaluation criteria (March & Smith, 1995; Hevner et al., 2004; Sonnenberg & vom Brocke, 2012). Two internal researchers and 11 AI experts across Germany were asked to execute the method according to its assessment process, in line with its categorization framework, using a representative selection of AI-Technologies identified from the literature (see Appendix IV). The participants were provided with copies of the MS Excel prototype pre-filled out with the identified AI-Technologies for use in the assessment. All the 11 AI experts came from different organizations, except two (see Table 8), but these two belonged to different departments of a Big Four. The DSR evaluation criteria were quantitatively and qualitatively measured through a questionnaire (see Appendix V) that captured the AI experts' experiences and opinions after they executed the AI-CC Method using the sample of AI-Technologies.

Number	Firm/Company/ Institution	Focus Area
1	Big Four	Audit Data Analytics
2	Big Four	Artificial Intelligence
3	University	Financial Mathematics
4	University	Audit Data Analytics
5	Second Tier	Certified Information Systems Auditor, Audit Data Analytics
6	Big Four	Forensics
7	Second Tier	Audit Data Analytics
8	Big Four	Artificial Intelligence
9	Independent	Forensics and Audit Data Analytics
10	Big Four	Artificial Intelligence
11	Second Tier	Audit Data Analytics

Table 8: AI Expert Details

6.2 Preliminary Evaluation Insights

To evaluate the effectiveness and efficiency of the developed artifact, two internal researchers independently assessed the initially selected AI-Technologies using the AI-CC Method. This setup is recognized in the literature because artifacts are commonly evaluated by two independent assessors (Wolfswinkel et al., 2013; Montazemi & Qahri-Saremi, 2015; Paré et al., 2015) and the two evaluators in the present study were both academic AI experts.

Activities 1–3 revealed that all the selected AI-Technologies (see Appendix IV) were assessable using the AI-CC Method because they coincided with the method's dimensions, requirements, and analytics components. However, Activities 4 and 5 revealed that it was not explicitly stated that the selected AI-Technologies, with their corresponding use cases from the literature, matched the functional and nonfunctional requirements. This left room for interpretation and discussion during the assessment of the AI-Technologies. When there were differences in assessment, the researchers discussed their respective interpretations and agreed on a common assessment for each requirement per AI-Technology. The combined assessment results of the two researchers were included in the provided MS Excel prototype under the tab "Internal_evaluation".

Furthermore, all the 11 AI experts executed the artifact's assessment process according to its categorization framework using the selected AI-Technologies and indicated their assessments in the MS Excel prototype. For each of the assessed AI-Technologies, the assessments made by all the 11 AI experts per requirement were aggregated and averaged to derive average assessment values. These results are illustrated in the MS Excel prototype under the tab "External_evaluation".

To ensure consistency between the researchers' and AI experts' assessments, we calculated the Cohen's Kappa value (Cohen, 1960) between the representative assessments.[9] The results ranged from 50 % to 92 %, with an average value of approximately 77 %, indicating substantial agreement (Landis & Koch, 1977). Figure 4 shows the Cohen's Kappa calculation for two AI-Technologies. Although there were some discrepancies in the assessment results of functional and nonfunctional requirements, the overall reliability measures of Cohen's Kappa between the internal and external evaluations confirmed the reliability of the AI-CC Method's assessment process.

AI-Technology with Use Case				Functional Requirements												Non-Functional Requirements													Cohen's Kappa
				Classes of Transactions (ISA 315.A190 (a) (Revised 2019))						Account Balances (ISA 315.A190 (b) (Revised 2019))						Quality-in-Use Model (ISO/IEC 25010:2011 and ISO/IEC 25019:2023)					Product Quality Model (ISO/IEC 25010:2023 and ISO/IEC 25010:2011)								
Study	Study Year	Authors	Evaluands	O	C	A	CO	CA	P	E	RO	C	AVA	CA	P	E1	E2	S	FR	CC	FS	PE	C	U	R	S	M	P	
Contract analytics in auditing	2019	Zhaokai, Y./Moffitt, K. C.	AI experts	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	83%
		Zhaokai, Y./Moffitt, K. C.	Researchers	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	
Prepare for takeoff: Improving asset measurement and audit quality with drone-enabled inventory audit procedures	2021	Christ, M. H. et al.	AI experts	1	1	1	0	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	90%
	2021	Christ, M. H. et al.	Researchers	1	1	1	0	0	0	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	

Figure 4: Cohen's Kappa

To summarize, the application of the AI-CC Method's assessment process 301 times[10] during the internal and external evaluations showed the effectiveness and efficiency of the AI-CC Method. Although Activities 4 and 5 seemed quite challenging for the internal

9 To calculate Cohen's Kappa, the following assumption was made: A threshold value was implemented to define if there is a fit for an AI-Technology with a specific requirement. If the researchers or AI experts assessed the applicability of an AI-Technology for a specific requirement as 3 or higher, we transformed the assessment to a value of 1 (as an indicator value for "applicable") for the specific requirement. If the assessment was below 3, we assigned a value of 0 (as an indicator value for "not applicable").

10 Each of the two researchers assessed 23 AI-Technologies, and all the 11 AI experts individually assessed 23 AI-Technologies. As discussed in the Demonstration section, two researchers also processed the artifact according to the deep autoencoder. This yielded to an overall iteration rate of 301 process instances.

researchers and the 11 AI experts, the reliability of the assessment process was confirmed. As AI is not yet robustly utilized in the auditing profession, and as there are different views on the potential of its use for auditing purposes, the broader range of agreement (50 % to 92 %) to the use of AI in the auditing profession should not raise concerns and is also likely to occur in real-world settings. Furthermore, as the AI-CC Method has an innovative design, initial difficulties in its execution are expected.

6.3 Overall Usefulness

After the execution of the AI-CC Method's assessment process, the 11 AI experts answered the questionnaire (see Appendix V) concerning the overall usefulness of the developed artifact. First, the respondents quantitatively rated the utility of the method on an ordinal scale based on their experiences with using the artifact.[11] The question regarding whether the AI-CC Method has utility in assessing AI-Technologies yielded an average score of 3,182 out of 5 points among all the 11 assessments received, which quantitatively confirms the AI-CC Method's utility.

Next, the overall usefulness of the artifact was measured more comprehensively in terms of novelty and importance, understandability and suitability, ease of use, operationality, robustness, and applicability and fidelity to real-world phenomena. In the following sections, the statements from the free-entry text fields in the completed questionnaires (see Appendix V) received from the 11 AI experts are analyzed.[12] These statements strengthen the relevance of the AI-CC Method as a DSR research contribution.

6.3.1 Novelty and Importance

The AI experts highlighted the *novelty* of the AI-CC Method as an assessment method, particularly in its approach to systematically assess AI-Technologies. One participant noted, "The novelty is high. I have not seen a systematic assessment of a *concrete* [emphasis added by AI-E 10] set of scientific publications [identified AI-Technologies] as to how they relate to auditing methods" (AI-E 10). Another participant said, "Use case-specific clustering, extensive literature review, and the finely structured evaluation measures [the requirements of the AI-CC Method] are the novel features of this assessment [method]" (AI-E 2). These statements suggest that the AI-CC Method is a unique approach to assessing AI-Technologies, emphasizing its comprehensive and structured nature. The breadth of AI-Technologies and the reliance on systematic judgment were also cited as novel aspects. A respondent said, "It is definitely good to have a broad range of applications of AI-Technologies that are evaluated" (AI-E 3). This indicates an appreciation of the AI-CC Method's ability to encompass a wide range of applications, although the challenge of possessing extensive knowledge in all these areas was recognized: "This requires very broad knowledge, which is hard to have for these [AI] technologies" (AI-E 3).

The *importance* of presenting specific AI-Technologies to AI experts, especially in the audit domain, was also emphasized: "It is important to present specific use cases [based

11 The ordinal scale is defined as follows: 0 = no utility; 1 = very low utility; 2 = low utility; 3 = utility; 4 = high utility; 5 = very high utility.

12 The grammatical errors in the direct quotes from the participants were corrected, but faithfulness to the original meaning of the text was ensured. Quotes received in German language, were translated. AI experts were abbreviated with "AI-E" and the corresponding number of the evaluation.

on AI-Technologies] to auditors. Abstract technologies are not readily assessable for the 'audit' application domain" (AI-E 4). The AI-CC Method is also comprehensive and has a forward-looking perspective, as confirmed by the following statement: "The developed assessment process is novel because it allows the analysis of different AI-Technologies and of the extent to which they can contribute to audit success from different perspectives" (AI-E 5).

Some AI experts confirmed the *importance* of the AI-CC Method in identifying technologies that can be used for auditing purposes: "[It] is important to identify the technologies that can improve, facilitate, or replace audit tasks" (AI-E 8) and "The importance of your assessment process lies in its ability to combine auditing assertions with canonical software quality criteria" (AI-E 11). This highlights the method's practical utility in enhancing audit processes.

However, one AI expert raised concerns about the alignment between the AI-CC Method's precision and the experimental nature of the identified AI-Technologies: "The clear-cut goal and preciseness of the assessment process do not match the content that the process tries to evaluate" (AI-E 6). It must be emphasized, that this concern depends heavily on the specific AI-Technology to be assessed. In future iterations, more detailed AI-Technologies with more technical details can be assessed. To address this concern, the AI-CC Method was demonstrated in technical detail based on deep autoencoders (see section 5).

6.3.2 Understandability and Suitability

Most of the AI experts who participated in the present study confirmed the *understandability* of the AI-CC Method, highlighting its clarity and comprehensiveness. The AI-CC Method was praised for its clear and comprehensive presentation, especially with the aid of the provided supplementary material (the MS Excel prototype). The following quotes illustrate this: "The assessment process is highly comprehensible" (AI-E 1); "The use case-specific [identified AI-Technologies] evaluation is very understandable" (AI-E 2); "The assessment process is understandable and easy to follow based on the given material" (AI-E 5); "The developed assessment process is easy to understand and suitable for its purpose" (AI-E 7); "The assessment process is easy to understand" (AI-E 8); "It [the assessment process] is well described" (AI-E 9); and "[the assessment process has] good understandability" (AI-E 10).

However, some participants (AI-E 3, AI-E 4, AI-E 11) had difficulties understanding the terminology used in the AI-CC Method, especially the nonfunctional requirements derived from the computer science community. One respondent said: "Auditors are comfortable with audit assertions, and the software evaluation criteria may enjoy widespread acknowledgment in 'techie' communities" (AI-E 11). This concern can be traced to the fact that AI has not been widely adopted by the auditing profession, and there are only few individuals who are experts in both AI and auditing. This was also experienced while trying to identify experts in both fields who could serve as evaluators in the present study. This study addressed this concern by including descriptions of the nonfunctional requirements in the Appendix III and the MS Excel prototype of the artifact provided as an electronic supplementary material.

As shown by the following quotes, the AI-CC Method's assessment process is valued for its relevance to and *suitability* for business and auditing matters: "The use case [identified

AI-Technologies]-specific evaluation is (…) suitable for business considerations" (AI-E 2) and "In my view, all important elements required for a suitability test for the relevant technologies are in place" (AI-E 1). However, one participant (AI-E 6) noted that the theoretical and untested nature of some AI-Technologies can lead to uncertainties in the assessment process, suggesting the inclusion of an "unknown" category in the AI-CC Method's terminology. It must be mentioned that this problem is based on the AI-Technologies that the AI experts were made to use in executing the AI-CC Method for assessment purposes. Providing a sample of AI-Technologies with a more empirically tested nature such as those from the computer science community, can address this concern. Therefore, AI-E 6's feedback does not directly affect the AI-CC Method's *understandability* and *suitability* and is more related to the sample of AI-Technologies being assessed. Furthermore, the "not assessable" category of the AI-CC Method's assessment process can take the place of the "unknown" category suggested to be included in the method's terminology.

6.3.3 Ease of Use, Operationality, and Robustness

The AI-CC Method's *ease of use* and *operationality* was confirmed: "In general, the assessment process is well guided and explained, and the dimensions are reasonable" (AI-E 6). The AI-CC Method was also commended for its comprehensive explanations, which were said to have enhanced its *ease of use*. The respondents appreciated the user-friendly features: "The comprehensive set of supplementary explanations made it [the method] easy to use" (AI-E 1); "The given material [AI-CC Method] was easy to use and good to operate" (AI-E 5); "The evaluation [assessment] process was easy, thanks to the structure" (AI-E 8); and "[The] ease of use/operationality was okay. I had no problem using the form [MS Excel prototype]" (AI-E 10).

The evaluation results for the DSR criterion of *operationality* point to an overall positive reception, with a few operational concerns. For instance, AI-E 9 criticized the MS Excel-based tool. It must be emphasized, though, that the current version of the AI-CC Method is in the first stage, involving presentation to the research community. Further development of the artifact could consider the adoption of a more user-friendly front end.

One AI expert raised other concerns: "While the dimensions evaluated are suitable in general, they do not fit many of the research papers [AI-Technologies] presented" (AI-E 6). However, as also mentioned in relation to the DSR criteria of *understandability* and *suitability*, this concern is not based on the artifact's design, but on the initially selected AI-Technologies. In further iterations of the artifact, AI experts will be made to assess more technically detailed AI-Technologies as illustrated in the demonstration section.

The *robustness* of the AI-CC Method was also confirmed: "The use case-specific [AI-Technology based] evaluation seems very robust and important for operational considerations" (AI-E 2). Several features of the AI-CC Method were commended by the AI experts for enhancing the method's robustness: "Its [the artifact's] robustness was greatly enhanced by the well understandable comments provided, and the use of drop-down menus [in the MS Excel prototype] effectively prevented erroneous or nonsensical entries" (AI-E 1). It was also stated that "the criteria covered most of the information regarding the papers [identified AI-Technologies] and were therefore robust" (AI-E 8). The following quote also shows the AI-CC Method's relevance as a central method of assessing AI-Technologies within an organization: "The [assessment] process was easy to communicate within the team and is also quite robust against differing opinions during the assessment stage" (AI-E 7).

However, a few AI experts expressed concerns regarding the method's *robustness*, focusing on the limited availability of experts in both the AI and risk-based auditing fields to conduct a comprehensive assessment (AI-E 3) and the method's limitations in meeting the operational needs of auditing (AI-E 4 and AI-E11). On this matter, it must be emphasized that the AI-CC Method aims to provide indications of the degree of applicability of AI-Technologies to risk-based auditing, rather than to provide concrete and final assessment values. In addition, research combining the fields of AI and risk-based auditing is not yet as vibrant as that of more traditional research areas. Thus, the experimental nature of some components of the developed AI-CC Method are to be expected and are also likely to be observed while developing the AI-CC Method for use in industrial settings.

6.3.4 Real-World Applicability and Fidelity

The AI-CC Method is viewed as having a strong potential to be applicable in real-world business contexts, particularly due to its coverage of relevant dimensions, requirements, and analytics and its structured assessment of AI-Technologies: "The relevant criteria on which a real-world suitability assessment (…) should be based are well covered by the assessment process" (AI-E 1); "Overall, the assessment process is applicable to real-world phenomena" (AI-E 5); and "Due to its real-world use case-specific evaluation of ML techniques, the assessment [AI-CC Method] is highly applicable in the business context" (AI-E 2).

However, it was also stated that the *real-world applicability* of the AI-CC Method depends highly on the AI-Technology to be assessed (AI-E 7). The AI-Technologies that the AI experts were asked to assess using the AI-CC Method were extracted from the literature. Thus, each of them was covered by a specific use case explained in the corresponding research paper identified (see Appendix IV for more details). Some of the AI-Technologies and their corresponding use cases provided a solid base for reliable assessments, but others, especially the AI-Technologies used for investigating accounting fraud, seemed relatively hard for a few AI experts to grasp (e.g., AI-E 4). To address this concern, a degree-of-applicability measure was added, providing a tendency-based value for identifying promising AI-Technologies; thus the method's design was adapted to the needs of AI experts, one of whom said: "I guess that kind of mean value or tendency of applicability and fidelity [of identified AI-Technologies] will be (…) good for the whole picture" (AI-E 3).

Challenges also exist regarding the artifact's *fidelity* to real-world phenomena. According to one respondent (AI-E 6), more details about the AI-Technologies must be made available by corresponding case studies with real-world settings to obtain more precise evaluation results and enhance the method's *fidelity* to real-world phenomena. Furthermore, some AI experts (AI-E 6, AI-E 9, AI-E 10) mentioned that the ideas stated regarding the selected AI-Technologies and their use cases were very promising but often lacked technical details to ensure a more precise assessment of the AI-Technologies' adherence to the functional and nonfunctional requirements. One respondent explained that "the more complex the ML solution, the more breaking and turning points there are in determining whether or not this approach can improve the overall audit delivery" (AI-E 6). The respondent concluded that "it usually takes many iterations and thorough testing to evaluate a solution approach" (AI-E 6). This indicates that the full extent of the complexity of assessing AI-Technologies cannot be considered in the initial development stage of the AI-CC Method. It will take several stages of development to further improve the artifact step by step. However, to mitigate the complexity problem, an AI-Technology Classification System was incorporated into the method after its evaluation. Furthermore, it was ensured that the AI-CC Method could always be updated to incorporate additional AI-Technologies (especially from the computer science community that had been investigated in more concrete case studies). Nevertheless, the AI-CC Method is a first step toward providing comprehensive guidance in the use of AI for the auditing profession, and its relevance for practice was confirmed by AI experts who participated in the present study during their evaluation of it, as shown by the following quote: "The assessment process [of the AI-CC Method] precisely captures the nature of ongoing debates about AI at our company"(AI-E 11).

7 Discussion

7.1 Overall Findings

The execution of the AI-CC Method in the present study using selected AI-Technologies provided general insights into the availability of promising AI-Technologies for gathering sufficient and appropriate audit evidence. The AI-Technologies used were selected from among the AI-Technologies identified by Appelbaum et al. (2018); the gathered insights can thus be seen as representative.[13]

As the external evaluation results illustrate, the calculated TVAs ranged from 48 to 79, with an average value of 66. High application values were derived from Text Mining approaches, such as contract analytics or the audit of confirmation letters with DL. Traditional ML approaches analyzing structured data, such as classification or clustering algorithms, also obtained high TVAs. Furthermore, ensemble learning based on boosting algorithms and Computer Vision with DL obtained high TVAs. Outside the scope of AI-Technologies, the evaluation showed that Process Mining is also associated with high TVAs, which may be attributable to the maturity and broader application of the technology within the auditing profession (Föhr, Reichelt, et al., 2023).

The low TVAs may be attributable to traditional ML approaches with NLP or Text-Mining use cases. As the evaluation results show, DL can analyze complex, unstructured data much better than traditional ML approaches can. Another reason for the low TVAs is the experimental character of NLP research contributions from the auditing and accounting community as they lack a link to specific functional and nonfunctional requirements. It must be mentioned that the NLP field has been undergoing a fundamental paradigm shift in recent months due to Large Language Models (LLMs) (Brown et al., 2020; Thoppilan et al., 2022). Thus, the implications of this for the auditing profession must be fully explored in the future. To evaluate LLMs according to the AI-CC Method, more research contributions in the auditing and accounting field are needed, such as work done by Eulerich & Wood (2023), Föhr, Schreyer, et al. (2023), and Gu et al. (2024).

The calculated DCS values ranged from 34 % to 55 %, with an average value of 42 %, indicating that most of the assessed AI-Technologies matched a broad range of require-

13 The following values refer to the MS Excel prototype's tab "External_evaluation" from the electronic supplementary material.

ments included in the AI-CC Method and therefore lacked a special-purpose perspective. This initially evaluated sample of AI-Technologies can be seen as only a starting point in the effort to further strengthen the discussion about the utility of AI-Technologies in gathering sufficient and appropriate audit evidence.

7.2 Limitations and Future Work

Research from the computer science community was not incorporated into the evaluation due to the innovative design of the artifact, which was likely to enhance its complexity for external evaluation procedures and lessen the chances of obtaining many responses as time efforts for evaluation would have significantly increased. Nevertheless, as discussed in the Demonstration section, the study demonstrated the adaptability of the AI-CC Method with more technically detailed and complex AI-Technologies demonstrating its components based on deep autoencoders for anomaly detection. Furthermore, the AI-CC Method can always be expanded so that it can be used to systematically assess other AI-Technologies.

The overarching goal in the initial development stage of the AI-CC Method was to obtain high-level indications of the applicability of several AI-Technologies to the auditing field. Indication-based assessment is highly relevant when it is not conducted by the one who originally developed the AI-Technology (vom Brocke et al., 2021). Assessment of a specific AI-Technology by its creator may be subject to severe biases, leading to the initial judgment or assessment of the AI-Technology as the most appropriate or best.

Further development stages of the AI-CC Method must address the remaining issues regarding the measured DSR evaluation criteria, such as going one step further from providing high-level indications of the applicability of AI-Technologies to providing granular and precise information about how these technologies can obtain audit evidence or cover risk assessment.

Nevertheless, the AI-CC Method proposed by the present study is a relevant DSR contribution, as confirmed by the results of its comprehensive evaluation. The study showed the innovative character of the developed AI-CC Method as the first systematic method of providing guidance for identifying AI-Technologies that can be used to gather sufficient and appropriate audit evidence, with corresponding implications for theory and practice.

8 Conclusion

There is an increasing need for the auditing profession to utilize AI-Technologies to achieve the primary objective of a risk-based audit, but there is no guidance regarding how to utilize these technologies in the auditing field. The present study sought to fill this gap proposing the AI-CC Method of identifying the most promising AI-Technologies for gathering sufficient and appropriate audit evidence and enhancing the effectiveness and efficiency of audit procedures. The study relied on DSR for its research design and used SME to develop the AI-CC Method's categorization framework and assessment process. The AI-CC Method meets the DSR contribution of "exaptation" because its initial assessment logic was based on vom Brocke et al. (2021), but it developed its own elements and components embedded within a completely different application context (Gregor & Hevner, 2013). The comprehensive evaluation of the artifact by two internal researchers and 11 AI experts across Germany confirmed its overall usefulness, effectiveness, and efficiency. The identified limitations of the developed artifact call for further development to promote the use of AI for auditing and accounting purposes.

Appendix I – Structured Literature Review Search

The present study followed the suggestion of Kogan et al. (2019) that the field of emerging technologies in auditing and accounting is best researched through the lens of Design Science Research (DSR). Thus, the study presented some selected research contributions (Webster & Watson, 2002), explicitly stating their adherence to DSR. The scope of the search terms utilized in the structured literature review was emerging technologies in general; thus, the search was broader than just focusing on AI-Technologies with corresponding Machine Learning (ML) and Deep Learning (DL) approaches. This more general approach was used because of the novelty and rarity of research contributions developed explicitly under the lens of DSR within the auditing and accounting profession.

The structured literature review had two main contributions: (1) it showed the current state of the art of DSR research explicitly applied in the auditing and accounting profession; and (2) it identified a research gap: that existing DSR in the auditing and accounting field lacks a comprehensive method for the systematic assessment of AI-Technologies under the risk-based audit approach.

The Digital Library of the American Accounting Association (AAA Digital Library) was the main database used in the structured literature review. This database includes the most relevant research papers for the scientific auditing and accounting profession. The initial search results were supplemented by the application of the search terms on the Web of Science Core Collection, Association for Computing Machinery (ACM) Digital Library, EconLit, Business Source Complete, ScienceDirect, and Wiley Online Library databases. The search terms "Design Science" and "Design Science Research" were specifically used for the AAA Digital Library due to the library's strong focus on the auditing and accounting profession. In addition, a thorough search was done on the AAA Digital Library by not restricting the search terms to titles, abstracts, keywords, or a specific publication timeframe[14]. For the Web of Science Core Collection, Business Source Complete, ACM Digital Library, EconLit, ScienceDirect, and Wiley Online Library databases, the more detailed and complex search term (("Design Science" OR "Design Science Research") AND ("Audit" OR "Auditing" OR "Accounting")) was used. To obtain only the targeted supplementary research papers in addition to the baseline search results from the AAA Digital Library, the search terms used in these databases were applied to titles, abstracts, and keywords. Furthermore, the searches on the supplementary databases were set within the publication timeframe of 01.01.2015 to 31.12.2022 so that only the most recent supplementary research papers would be extracted, in addition to the baseline sample extracted from the AAA Digital Library. The search process was conducted in sequential order (vom Brocke et al., 2015). Applying the search terms to the aforementioned databases, led to 505 hits. After a review of the titles and abstracts of these 505 articles and the application of the inclusion criteria regarding the utilization of emerging technologies and their links to risk-based audit procedures (ISA 315 (Revised 2019), IAASB, 2019 and ISA 330, IFAC, 2009c), the articles most relevant to the study's research purpose remained.

14 The last publication date of research papers was set to 31.12.2022.

Appendix II – AI-Technology Classification System

Data Structure

Category	Description
Structured data	Structured data are data formats stored in relational databases for efficient and direct computation through AI-Technologies (Lee, 2017).
Unstructured data	Unstructured data are data formats without the required structural definition for direct processing by AI-Technologies, (e.g., text, audio, and video data) (Gandomi & Haider, 2015).

Table A-1: Data Structure

Domain

Procedure	Description
Data Mining	Data Mining involves identifying meaningful patterns, models, and insights within vast amounts of data (Han et al., 2022).
NLP	Natural Language Processing (NLP) focuses on the understanding and utilization of human language through AI-Technologies to perform specific tasks (Ghosh & Gunning, 2019; Chowdhary, 2020).
Text Mining	Text Mining analyses unstructured data in the form of text data in document collections, for example, to extract unknown patterns in unstructured data (Feldman & Sanger, 2007; Tseng et al., 2007).
Process Mining	Process Mining is about "to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's systems" (van der Aalst, 2016, p. 31).
Computer Vision	Computer Vision aims to describe the world that humans perceive and to reconstruct the properties of the environment (Szeliski, 2022). Thus, it extracts information from image or video data to provide judgment support by, for example, classifying or detecting objects in visual data (International Business Machines Corporation, 2024).

Table A-2: Domain

Paradigm

Learning Paradigm	Description
Supervised	In supervised Machine Learning (ML) the algorithm is given input-output pairings to learn the underlying function and to map attribute values to the target attribute (also called label) (Kelleher & Tierney, 2018; Russell & Norvig, 2022).
Unsupervised	Unsupervised ML can be broadly defined as learning without feedback. Thus, it does not define a target variable (Han et al., 2022; Kelleher & Tierney, 2018).

Table A-3: Paradigm

Method

Method for Problem Solving	Definition
Classification	The function of classification is to separate data objects by distinguishing a limited set of unique values (Russell & Norvig, 2022).
Regression	The outputs of regression are continuous values for the estimation of expectations (Bertomeu et al., 2021).
Clustering	Clustering describes grouping into mutually similar data points based on their attribute values and utilized distance measures (Jain et al., 1999).
Anomaly/Outlier detection	An outlier can be seen as a data point with a significant deviation from the other data points regarding the corresponding values of their attributes. One can assume the generation of an outlier by another model in comparison to the other data objects (Han et al., 2022).
Decision support	Decision support is based on the underlying task specified by its user. It aims to support the cognitive process of individuals and ideally yields a decision action that has not occurred without the specific decision support (Keen, 1980).
Process discovery, enhancement and conformance checking (Process visualization)	Process discovery refers to the creation of a process model based on event logs. Process enhancement extends the control flow perspective derived by process discovery through additional perspectives (e.g., organizational, time). Conformance checking can be defined by comparing a "to-be" process model with an "as-is" process model to find similarities and material deviations (van der Aalst, 2016).
Prediction	Prediction can be defined as the prediction of future outcomes based on input data through ML algorithms. The resulting model output is often measured by the accuracy of the out-of-sample performance (Saeedi, 2021).
Object detection	Object detection refers to the detection and identification of instances of target objects in image data through a classification model (Amit et al., 2021; Christ et al., 2021).

Table A-4: Method

Approach

Category	Approach	Description	References
Classification algorithms	Decision Trees	Decision Trees are classifiers based on a tree-like structure where each node represents a test for the dataset with corresponding attribute values, each branch represents an outcome of the test, and each leaf node corresponds to an appropriate class label. The beginning of a decision tree can be determined by the root node.	Quinlan (1986); Wu et al. (2008)
	Naïve Bayes	Naïve Bayes is a probabilistic classifier based on probability measures. A given tuple can be allocated to a specific class label. One main basic component of the Naïve Bayes Classifier is the Bayes' Theorem.	Rish (2001); Wu et al. (2008); Han et al. (2022)
	Support Vector Machines	Support Vector Machines (SVM) are used for classifying linear and non-linear data. Its main function is the construction of a hyperplane based on training data to geometrically maximize the margin for the separation of different classes.	Vapnik (2000); Wu et al. (2008); Ge et al. (2017)
	Hidden-Markov-Model	In the context of NLP, Hidden Markov Models (HMMs) are statistical models used to predict sequences of hidden states based on observable sequences. They are often employed for tasks such as part-of-speech tagging, where the observable sequence is a series of words and the hidden states represent the grammatical categories (e.g., noun, verb) for each word.	Rabiner & Juang (1986); Kouemou (2011); Raschke et al. (2018)
Clustering algorithms	k-means algorithm	The k-means algorithm is a clustering algorithm that randomly selects data points as centroids representing cluster means. These cluster means are iteratively refined until there is no longer any improvement in the within-cluster variation.	Hartigan (1975); Bishop (2006)
	DBSCAN	Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an unsupervised learning algorithm for handling the problem of finding clusters of arbitrary shapes. The algorithm is based on a threshold for the minimum number of necessary neighbors (MinPts) within a clearly defined radius ε.	Ester et al. (1996); Schubert et al. (2017)
	Complete Link Hierarchical Algorithm	Complete Link Hierarchical Clustering can be defined as agglomerative clustering. Dissimilarity between two cluster groups is measured by the farthest neighbor distance.	Tan et al. (2019); Gan et al. (2020)

Category	Approach	Description	References
Artificial Neural Networks	Feedforward Neural Network	The basic structure of an Artificial Neural Network (ANN) is determined by an input layer for obtaining the initial data attributes, a hidden layer for weighting and computing more abstract attribute features, and an output layer for obtaining the output of the target value. The smallest elements of the ANN form neurons with weighted connections to the upstream and downstream layers.	Mitchell (1997); LeCun et al. (2015)
Artificial Neural Networks	Convolutional Neural Network (CNN)	Convolutional Neural Networks (CNN) are often utilized in Computer Vision tasks. Instead of fully connected input, hidden, and output layers, a CNN is implemented with the three main components: a convolutional layer, a pooling layer, and a fully connected layer.	LeCun et al. (1989); LeCun et al. (2015)
Ensemble Learning	RUSBoost	RUSBoost can be defined as a variant of AdaBoost utilizing the general functionality of Boosting and additionally addressing class imbalance learning problems. Based on several weak classifiers, a strong classifier is derived. Therefore, RUSBoost uses random undersampling by utilizing the full sample of the minority class and a randomly generated subsample of the majority class regarding each training iteration for model building.	Freund & Schapire (1996, 1997); Schapire & Singer (1999); Seiffert et al. (2010)
	Gradient Boosted Regression Tree	The Gradient Boosted Regression Tree (GBRT) handles regression problems. GBRT initially splits the dataset into two subgroups; these subgroups are iteratively split again to improve performance accuracy. Within each split, the resulting residuals are computed. Based on these residuals, further decision trees are created.	Friedman (2001)
Process Mining	Alpha Algorithm	The alpha algorithm is one of the first algorithms to perform process discovery. The algorithm considers the relationships between individual activities based on event logs.	van der Aalst et al. (2004); van der Aalst (2010, 2016)
	Fuzzy Miner	The fuzzy miner algorithm enables complexity reduction when performing process discovery techniques by focusing on the most essential process sequences.	Günther and van der Aalst (2007); Jans et al. (2014)

Table A-5: Approach

Appendix III – Glossary of Requirements

Functional Requirements

Classes of Transactions	Definition	Source
Occurrence (O)	Transactions and events that have been stated in the financial statements occurred and are attributable to the entity.	ISA 315.A190 (a) (i) (Revied 2019)
Completeness (C)	All relevant transactions and events have been recorded, and all relevant disclosures have been included.	ISA 315.A190 (a) (ii) (Revied 2019)
Accuracy (A)	Amounts or other data of transactions and events have been recorded appropriately, and disclosures have been measured and described appropriately.	ISA 315.A190 (a) (iii) (Revied 2019)
Cutoff (CO)	Transactions and events have been assigned to the correct period.	ISA 315.A190 (a) (iv) (Revied 2019)
Classification (CA)	Transactions and events have been assigned to the correct accounts.	ISA 315.A190 (a) (v) (Revied 2019)
Presentation (P)	Transactions and events are appropriately summed or broken-down and transparently described, and disclosures are significant and comprehensible regarding the requirements of the applicable financial reporting framework.	ISA 315.A190 (a) (vi) (Revied 2019)

Table A-6: Part I of Functional Requirements defined by ISA 315.A190 (a) (Revised 2019), (IAASB, 2019)

Account Balances	Definition	Source
Existence (E)	The existence of assets, liabilities, and equity interests.	ISA 315.A190 (b) (i) (Revied 2019)
Rights and obligations (RO)	The rights over the assets and the obligations resulting from the liabilities are attributable to the entity.	ISA 315.A190 (b) (ii) (Revied 2019)
Completeness (C)	All relevant assets, liabilities and equity interests have been recorded, and all relevant disclosures have been included.	ISA 315.A190 (b) (iii) (Revied 2019)
Accuracy, valuation, and allocation (AVA)	Assets, liabilities, and equity interests have been included in the financial statements at the correct amounts and relevant adjustments (valuation or allocation) have been appropriately made. Related disclosures have been correctly measured and described.	ISA 315.A190 (b) (iv) (Revied 2019)

Account Balances	Definition	Source
Classification (CA)	Assets, liabilities, and equity interests have been assigned to the correct accounts.	ISA 315.A190 (b) (v) (Revied 2019)
Presentation (P)	Assets, liabilities, and equity interests are appropriately summed or broken-down and transparently described, and disclosures are significant and comprehensible regarding the requirements of the applicable financial reporting framework.	ISA 315.A190 (b) (vi) (Revied 2019)

Table A-7: Part II of Functional Requirements defined by ISA 315.A190 (b) (Revised 2019), (IAASB, 2019)

Nonfunctional Requirements

Quality-in-Use Model	Definition	Source
Effectiveness (E1)	Degree of accuracy and completeness of the AI-Technology to meet user specified goals.	ISO/IEC 25010:2011, 4.1.1
Efficiency (E2)	Resource consumption in relation to effectiveness by the application of the AI-Technology.	ISO/IEC 25010:2011, 4.1.2
Satisfaction (S)	Degree to which user specified goals are achieved by utilizing the AI-Technology.	ISO/IEC 25010:2011, 4.1.3
Freedom from risk (FR)	Degree to which the AI-Technology alleviates potential risk factors, regarding economy, human, health, and environment.	ISO/IEC 25010:2011, 4.1.4 and ISO/IEC 25019:2023, 3.2.2
Context coverage (CC)	Degree to which the AI-Technology can be utilized regarding effectiveness, efficiency, satisfaction, and freedom from risk in different contexts.	ISO/IEC 25010:2011, 4.1.5

Table A-8: Quality-in-Use Model (ISO/IEC 25010:2011, ISO/IEC, 2011 and ISO/IEC 25019:2023, ISO/IEC, 2023b)

Product Quality Model	Definition	Source
Functional suitability (FS)	Degree to which the functionalities of the AI-Technology meet stated and implied needs regarding specific conditions.	ISO/IEC 25010:2023, 3.1
Performance efficiency (PE)	Performance of the AI-Technology in relation to the number of resources utilized under specific conditions.	ISO/IEC 25010:2023, 3.2
Compatibility (C)	Degree to which the AI-Technology can exchange outputs with other algorithms, and/or execute the required functions sharing the same system components.	ISO/IEC 25010:2023, 3.3
Usability (U)	Degree to which the AI-Technology can be used by specific users to achieve specified goals regarding effectiveness, efficiency, and satisfaction in a specified context.	ISO/IEC 25010:2011, 4.2.4
Reliability (R)	Degree to which the AI-Technology fulfills its specified functionality regarding specified conditions and a specified period with no failures.	ISO/IEC 25010:2023, 3.5
Security (S)	Degree to which the AI-Technology protects information and data regarding the hierarchical policy of authorization.	ISO/IEC 25010:2023, 3.6
Maintainability (M)	Degree of effectiveness and efficiency with which the AI-Technology can be corrected, improved, or adapted by the intended maintainers.	ISO/IEC 25010:2023, 3.7
Portability (P)	Degree of effectiveness and efficiency with which the AI-Technology can be transferred from one hardware, software, and environment component to other ones.	ISO/IEC 25010:2011, 4.2.8

Table A-9: Product Quality Model (ISO/IEC 25010:2023, ISO/IEC, 2023a and ISO/IEC 25010:2011, ISO/IEC, 2011)

Appendix IV – Structured AI-Technology Identification

To conduct an evaluation of the developed artifact and in view of the use case-specific nature of AI-Technologies, the study identified AI-Technologies in the context of a risk-based audit. Therefore, a structured literature review (vom Brocke et al., 2009; vom Brocke et al., 2015) was conducted on the AAA Digital Library, Web of Science Core Collection, EconLit, and Business Source Complete databases. Documenting the search process meets the rigor requirements of DSR (Hevner et al., 2004). The timeframe was set from 01.01.2015 to 31.01.2022. The search terms below, relying on representative AI-Technologies, were used to obtain relevant use cases.

- ("Support Vector Machines" AND (Audit OR Auditing))
- (Clustering AND (Audit OR Auditing))
- ("Neural Networks" AND (Audit OR Auditing))
- ("Natural Language Processing" AND (Audit OR Auditing))
- ("Ensemble Learning" AND (Audit OR Auditing))
- ("Process Mining" AND (Audit OR Auditing))
- (Drones AND (Audit OR Auditing))

To obtain highly valued evaluation results and to avoid overly extending the evaluation process, only one to three use cases for each stated AI-Technology were extracted. It is not claimed that the study completely extracted all relevant AI-Technologies, but it is maintained that the selection is representative and well suited for evaluation purposes. However, the developed artifact is arbitrarily expandable with further AI-Technologies. Shown in the table below are the extracted AI-Technologies with use cases.

AI-Technology	Use Case
Support Vector Machines	**Saeedi (2021):** Audit opinion prediction: A comparison of data mining techniques
	Nasir et al. (2021): Developing a decision support system to detect material weaknesses in internal control
	Chen et al. (2017): Enhancement of fraud detection for narratives in annual reports
Clustering	**No et al. (2019):** Multidimensional audit data selection (MADS): A framework for using data analytics in the audit data selection process
	Byrnes (2019): Automated clustering for data analytics
	Yan et al. (2019): Research on application of data mining technology in risk assessment process of audit
Text Mining	**Sun (2019):** Applying deep learning to audit procedures: An illustrative framework (Use cases: Internal control evaluation, substantive test, and completion)

AI-Technology	Use Case
Natural Language Processing	**Sun (2019):** Applying deep learning to audit procedures: An illustrative framework (Use case: Audit planning)
	Burns and Igou (2019): "Alexa, write an audit opinion": Adopting intelligent virtual assistants in accounting workplaces
	Li and Liu (2020): Development of an intelligent NLP-based audit plan knowledge discovery system
Natural Language Processing	**Raschke et al. (2018):** AI-enhanced audit inquiry: A research note
	Zhaokai and Moffitt (2019): Contract analytics in auditing
Ensemble Learning	**Hooda et al. (2020):** Optimizing fraudulent firm prediction using ensemble machine learning: A case study of an external audit
	Bao et al. (2020): Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach
	Bertomeu et al. (2021): Using machine learning to detect misstatements
Process Mining	**Chiu and Jans (2019):** Process mining of event logs: A case study evaluating internal control effectiveness
	Wang et al. (2020): Redesigning business process to comply with the new revenue recognition standard using process mining
	Werner et al. (2021): Embedding process mining into financial statement audits
Computer Vision	**X. Liu et al. (2021):** Automatic detection of oil palm tree from UAV images based on the deep learning method
	Christ et al. (2021): Prepare for takeoff: Improving asset measurement and audit quality with drone-enabled inventory audit procedures
	Appelbaum & Nehmer (2017): Using drones in internal and external audits: An exploratory framework

Table A-10: Sample of AI-Technologies

Appendix V – Questionnaire

Do you think that the assessment process of the AI-CC Method has good utility in assessing AI-Technologies' adherence to functional and nonfunctional requirements? Please assess the utility based on the scale below (whereas 0 = no utility; and 5 = very high utility).

0 ☐
1 ☐
2 ☐
3 ☐
4 ☐
5 ☐

Please evaluate, in your own words, the novelty and importance of the developed AI-CC Method. Please write your answer in the box below.

Please evaluate, in your own words, the understandability and suitability of the developed AI-CC Method. Please write your answer in the box below.

Please evaluate, in your own words, the ease of use, operationality, and robustness of the developed AI-CC Method. Please write your answer in the box below.

Please evaluate, in your own words, the developed AI-CC Method's applicability to and fidelity to real-world phenomena. Please write your answer in the box below.

References

Albizri, A., & Appelbaum, D. (2021). Trust but verify: The oracle paradox of blockchain smart contracts. *Journal of Information Systems*, *35*(2), 1–16. https://doi.org/10.2308/ISYS-19-024

Alzamil, Z. S., Appelbaum, D., Glasgall, W., & Vasarhelyi, M. A. (2021). Applications of data analytics: Cluster analysis of not-for-profit data. *Journal of Information Systems*, *35*(3), 199–221. https://doi.org/10.2308/ISYS-2020-025

Amit, Y., Felzenszwalb, P., & Girshick, R. (2021). Object detection. In K. Ikeuchi (Ed.), *Computer vision* (pp. 875–883). Springer. https://doi.org/10.1007/978-3-030-63416-2_660

Appelbaum, D., & Nehmer, R. A. (2017). Using drones in internal and external audits: An exploratory framework. *Journal of Emerging Technologies in Accounting*, *14*(1), 99–113. https://doi.org/10.2308/jeta-51704

Appelbaum, D., & Nehmer, R. A. (2020). Auditing cloud-based blockchain accounting systems. *Journal of Information Systems*, *34*(2), 5–21. https://doi.org/10.2308/isys-52660

Appelbaum, D. A., Kogan, A., & Vasarhelyi, M. A. (2018). Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics. *Journal of Accounting Literature*, *40*(1), 83–101. https://doi.org/10.1016/j.acclit.2018.01.001

Austin, A. A., Carpenter, T. D., Christ, M. H., & Nielson, C. S. (2021). The data analytics journey: Interactions among auditors, managers, regulation, and technology. *Contemporary Accounting Research*, *38*(3), 1888–1924. https://doi.org/10.1111/1911-3846.12680

Avison, D. E. (1996). Information systems development methodologies: A broader perspective. In S. Brinkkemper, K. Lyytinen, & R. J. Welke (Eds.), *Method engineering. ME 1996. IFIP – The International Federation for Information Processing* (pp. 263–277). Springer. https://doi.org/10.1007/978-0-387-35080-6_17

Azimi, M., & Agrawal, A. (2021). Is positive sentiment in corporate annual reports informative? Evidence from deep learning. *The Review of Asset Pricing Studies*, *11*(4), 762–805. https://doi.org/10.1093/rapstu/raab005

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Bennis, W., & O'Toole, J. (2005). How business schools lost their way. *Harvard Business Review*, *83*(5), 96–104.

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2021). Using machine learning to detect misstatements. *Review of Accounting Studies*, *26*, 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Boritz, J. E., & No, W. G. (2016). Computer-assisted functions for auditing XBRL-related documents. *Journal of Emerging Technologies in Accounting*, *13*(1), 53–83. https://doi.org/10.2308/jeta-51436

Braun, C., Wortmann, F., Hafner, M., & Winter, R. (2005). Method construction: A core approach to organizational engineering. *Proceedings of the 2005 ACM symposium on applied computing*, 1295–1299. https://doi.org/10.1145/1066677.1066971

Brenner, W., Van Giffen, B., & Koehler, J. (2020). Management of artificial intelligence: Feasibility, desirability and viability. In S. Aier, P. Rohner, & J. Schelp (Eds.), *Engineering the transformation of the enterprise* (pp. 15–36). Springer. https://doi.org/10.1007/978-3-030-84655-8_2

Brinkkemper, S. (1996). Method engineering: engineering of information systems development methods and tools. *Information and Software Technology*, *38*(4), 275–280. https://doi.org/10.1016/0950-5849(95)01059-9

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Infromation Processing*, *33*, 1877–1901.

Brunner, M., Sillaber, C., Demetz, L., Manhart, M., & Breu, R. (2018). Towards data-driven decision support for organizational IT security audits. *Information Technology*, *60*(4), 207–217. https://doi.org/10.1515/itit-2018-0002

Burns, M. B., & Igou, A. (2019). "Alexa, write an audit opinion": Adopting intelligent virtual assistants in accounting workplaces. *Journal of Emerging Technologies in Accounting*, *16*(1), 81–92. https://doi.org/10.2308/jeta-52424

Burton, F. G., Heninger, W. G., Summers, S. L., & Wood, D. A. (2022). Updated perceptions of accounting academics on the review and publication process. *SSRN Working Paper*. https://dx.doi.org/10.2139/ssrn.3809906

Burton, F. G., Summers, S. L., Wilks, T. J., & Wood, D. A. (2021). Do we matter? Attention the general public, policymakers, and academics give to accounting research. *Issues in Accounting Education 36*(1), 1–22. https://doi.org/10.2308/ISSUES-2020-017

Burton, F. G., Summers, S. L., Wilks, T. J., & Wood, D. A. (2022). Relevance of accounting research (ROAR) scores: Ratings of titles and abstracts by accounting professionals. *Accounting Horizons*, *36*(2), 7–18. https://doi.org/10.2308/HORIZONS-2020-147

Byrnes, P. E. (2019). Automated clustering for data analytics. *Journal of Emerging Technologies in Accounting*, *16*(2), 43–58. https://doi.org/10.2308/jeta-52474

Centobelli, P., Cerchione, R., Del Vecchio, P., Oropallo, E., & Secundo, G. (2022). Blockchain technology design in accounting: Game changer to tackle fraud or technological fairy tale? *Accounting, Auditing & Accountability Journal*, *35*(7), 1566–1597. https://doi.org/10.1108/AAAJ-10-2020-4994

Chen, Y.-J., Wu, C.-H., Chen, Y.-M., Li, H.-Y., & Chen, H.-K. (2017). Enhancement of fraud detection for narratives in annual reports. *International Journal of Accounting Information Systems*, *26*, 32–45. https://doi.org/10.1016/j.accinf.2017.06.004

Chiu, T., & Jans, M. (2019). Process Mining of event logs: A case study evaluating internal control effectiveness. *Accounting Horizons*, *33*(3), 141–156. https://doi.org/10.2308/acch-52458

Chou, C.-C., Chang, C. J., Chin, C.-L., & Chiang, W.-T. (2018). Measuring the consistency of quantitative and qualitative information in financial reports: A design science approach. *Journal of Emerging Technologies in Accounting*, *15*(2), 93–109. https://doi.org/10.2308/jeta-52312

Chou, C.-C., Chang, C. J., & Peng, J. (2016). Integrating XBRL data with textual information in Chinese: A semantic web approach. *International Journal of Accounting Information Systems*, *21*, 32–46. https://doi.org/10.1016/j.accinf.2016.04.002

Chowdhary, K. R. (2020). *Fundamentals of Artificial intelligence*. Springer. https://doi.org/10.1007/978-81-322-3972-7

Christ, M. H., Emett, S. A., Summers, S. L., & Wood, D. A. (2021). Prepare for takeoff: Improving asset measurement and audit quality with drone-enabled inventory audit procedures. *Review of Accounting Studies*, *26*(4), 1323–1343. https://doi.org/10.1007/s11142-020-09574-5

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *The Academy of Management Review*, *36*(1), 12–32. https://doi.org/10.5465/amr.2009.0486

Coss, D. L., & Dhillon, G. (2020). A framework for auditing and strategizing to ensure cloud privacy. *Journal of Information Systems*, *34*(2), 47–63. https://doi.org/10.2308/isys-17-046

Denner, M.-S., Püschel, L., & Röglinger, M. (2018). How to exploit the digitalization potential of business processes. *Business and Information Systems Engineering*, *60*(4), 1–19. https://doi.org/10.1007/s12599-017-0509-x

Deußer, T., Pielka, M., Pucknat, L., Jacob, B., Dilmaghani, T., Nourimand, M., Kliem, B., Loitz, R., Bauckhage, C., & Sifa, R. (2023). Contradiction detection in financial reports. *Proceedings of the Nothern Lights Deep Learning Workshop*, *4*. https://doi.org/10.7557/18.6799

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (KDD'96), 226–231. https://cdn.aaai.org/KDD/1996/KDD96-037.pdf

Eulerich, M., Pawlowski, J., Waddoups, N. J., & Wood, D. A. (2022). A framework for using robotic process automation for audit tasks. *Contemporary Accounting Research*, *39*(1), 691–720. https://doi.org/10.1111/1911-3846.12723

Eulerich, M., & Wood, D. A. (2023). A demonstration of how ChatGPT can be used in the internal auditing process. *SSRN Working Paper*. https://dx.doi.org/10.2139/ssrn.4519583

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies*, *27*, 938–985. https://doi.org/10.1007/s11142-022-09697-x

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.

Feliciano, C., & Quick, R. (2022). Innovative information technology in auditing: Auditor's perceptions of future importance and current auditor expertise. *Accounting in Europe*, *19*(2), 311–331. https://doi.org/10.1080/17449480.2022.2046283

Föhr, T. L., Reichelt, V., Marten, K.-U., & Eulerich, M. (2023). A framework for the structured implementation of process mining for audit tasks. *SSRN Working Paper*. https://dx.doi.org/10.2139/ssrn.4478182

Föhr, T. L., Schreyer, M., Juppe, T. A., & Marten, K.-U. (2023). Assuring sustainable futures: Auditing sustainability reports using AI foundation models. *SSRN Working Paper*. https://dx.doi.org/10.2139/ssrn.4502549

Freiman, J. W., Kim, Y., & Vasarhelyi, M. A. (2022). Full population testing: Applying multidimensional audit data sampling (MADS) to general ledger data auditing. *International Journal of Accounting Information Systems*, *46*, Article 100573. https://doi.org/10.1016/j.accinf.2022.100573

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning (ICML'96)*, 148–156. https://dl.acm.org/doi/10.5555/3091696.3091715

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The annals of Statistics*, *29*(5), 1189–1232.

Gan, G., Ma, C., & Wu, J. (2020). *Data clustering: Theory, algorithms, and applications* (2nd ed.). Society for Industrial and Applied Mathematics.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gartner. (2024, March 15). *What is artificial intelligence?* https://www.gartner.com/en/topics/artificial-intelligence

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, *5*, 20590–20616. https://doi.org/10.1109/ACCESS. 2017.2756872

Ghosh, S., & Gunning, D. (2019). *Natural language processing fundamentals: Build intelligent applications that can interpret the human language to deliver impactful results*. Packt Publishing Ltd.

Gierbl, A., Schreyer, M., Borth, D., & Leibfried, P. (2021). Deep Learning für die Wirtschaftsprüfung: Eine Darstellung von Theorie, Funktionsweise und Anwendungsmöglichkeiten. *Zeitschrift für internationale Rechnungslegung (IRZ)*, *16*(7/8), 317–322.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312–335. https://doi.org/10.17705/1jais.00129

Gu, H., Schreyer, M., Moffitt, K., & Vasarhelyi, M. A. (2024). Artificial intelligence co-piloted auditing. *International Journal of Accounting Information Systems*, *54*, Article 100698. https://doi.org/10.1016/j.accinf.2024.100698

Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy mining – Adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *Business process management. BPM 2007. Lecture notes in computer science* (Vol. 4714, pp. 328–343). Springer. https://doi.org/10.1007/978-3-540-75183-0_24

Guo, K. H., Yu, X., & Wilkin, C. (2022). A picture is worth a thousand journal entries: Accounting graph typology for auditing and fraud detection. *Journal of Information Systems*, *36*(2), 53–81. https://doi.org/10.2308/ISYS-2021-003

Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Hartigan, J. A. (1975). *Clustering Algorithms*. John Wiley & Sons.

Hawkins, S., He, H., Williams, G., & Baxter, R. (2002). Outlier detection using replicator neural networks. In Y. Kambayashi, W. Winiwarter, & M. Arikawa (Eds.), *Data warehousing and knowledge discovery. DaWak 2002. Lecture notes in computer science* (Vol. 2454, pp. 170–180). Springer. https://doi.org/10.1007/3-540-46145-0_17

Henderson-Sellers, B., & Ralyté, J. (2010). Situational method engineering: State-of-the-art review. *Journal of Universal Computer Science*, *16*(3), 424–478.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hillebrand, L., Deußer, T., Dilmaghani, T., Kliem, B., Loitz, R., Bauckhage, C., & Sifa, R. (2022). KPI-BERT: A joint named entity recognition and relation extraction model for financial reports. *2022 26th International Conference on Pattern Recognition (ICPR)*, 606–612. https://doi.org/10.1109/ICPR56361.2022.9956191

Hooda, N., Bawa, S., & Rana, P. S. (2020). Optimizing fraudulent firm prediction using ensemble machine learning: A case study of an external audit. *Applied Artificial Intelligence*, *34*(1), 20–30. https://doi.org/10.1080/08839514.2019.1680182

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Huang, S.-M., Wang, T. D., Yen, J.-C., Lee, C.-B., Wang, Y.-C., & Yang, Y.-T. (2022). The use of geographic information in audit data analytics for evidence gathering: A design science approach. *Journal of Information Systems*, *36*(3), 115–128. https://doi.org/10.2308/ISYS-2020-045

International Auditing and Assurance Standards Board (2019). *International Strandard on Auditing 315 (Revised 2019): Identifying and assessing the risks of material misstatement*. Retrieved March 16, 2024 from https://www.iaasb.org/publications/isa-315-revised-2019-identifying-and-assessing-risks-material-misstatement

International Auditing and Assurance Standards Board. (2022). *ISA 315 First-time implementation guide*. Retrieved March 16, 2024 from https://www.iaasb.org/publications/isa-315-first-time-implementation-guide

International Auditing and Assurance Standards Board (2024, March 16). *Technology*. https://www.iaasb.org/focus-areas/technology

International Business Machines Corporation (2024, March 16). *What is computer vision?* https://www.ibm.com/topics/computer-vision

International Federation of Accountants (2009a). *International Standard on Auditing 200: Overall objectives of the independent auditor and the conduct of an audit in accordance with international standards on auditing*. Retrieved March 16, 2024 from https://www.ifac.org/_flysystem/azure-private/publications/files/A009%202012%20IAASB%20Handbook%20ISA%20200.pdf

International Federation of Accountants (2009b). *International Standard on Auditing 240: The auditor's responsibilities relating to fraud in an audit of financial statements*. Retrieved March 16, 2024 from https://www.ifac.org/_flysystem/azure-private/publications/files/A012%202013%20IAASB%20Handbook%20ISA%20240.pdf

International Federation of Accountants (2009c). *International Standard on Auditing 330: The auditor's responses to assessed risks*. Retrieved March 16, 2024 from https://www.ifac.org/_flysystem/azure-private/publications/files/A019%202013%20IAASB%20Handbook%20ISA%20330.pdf

International Federation of Accountants (2009d). *International Standard on Auditing 520: Analytical procedures*. Retrieved March 16, 2024 from https://www.ifac.org/_flysystem/azure-private/publications/files/A026%202013%20IAASB%20Handbook%20ISA%20520.pdf

International Organization for Standardization / International Electrotechnical Commission (2011). *ISO/IEC 25010:2011: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models*. Retrieved March 16, 2024 from https://www.iso.org/standard/35733.html

International Organization for Standardization / International Electrotechnical Commission (2023a). *ISO/IEC 25010:2023: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Product quality model*. Retrieved March 16, 2024 from https://www.iso.org/standard/78176.html

International Organization for Standardization / International Electrotechnical Commission (2023b). *ISO/IEC 25019:2023: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality-in-use model*. Retrieved March 16, 2024 from https://www.iso.org/standard/78177.html

International Organization for Standardization / International Electrotechnical Commission (2024). *ISO/IEC 25002:2024: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality model overview and usage*. Retrieved March 16, 2024 from https://www.iso.org/standard/78175.html

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), 1–20. https://doi.org/10.2308/jeta-10511

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review*, *89*(5), 1751–1773. https://doi.org/10.2308/accr-50807

Keen, P. G. W. (1980). Decision support systems: A research perspective. In G. Fick & R. H. Sprague (Eds.), *Decision Support Systems: Issues and Challenges* (pp. 23–44). Pergamon Press.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. The MIT Press.

Kiesow, A., Zarvić, N., & Thomas, O. (2015). Design science for future AIS: Transferring continuous auditing issues to a gradual methodology. In B. Donnellan, M. Helfert, J. Kenneally, D. VanderMeer, M. Rothenberger, & R. Winter (Eds.), *New horizons in design science: Broadening the research agenda. DESRIST 2015. Lecutre notes in computer vision* (Vol. 9073, pp. 311–326). Springer. https://doi.org/10.1007/978-3-319-18714-3_20

Kogan, A., Mayhew, B. W., & Vasarhelyi, M. A. (2019). Audit data analytics research — An application of design science methodology. *Accounting Horizons*, *33*(3), 69–73. https://doi.org/10.2308/acch-52459

Kogan, A., & Yin, C. (2021). Privacy-preserving information sharing within an audit firm. *Journal of Information Systems*, *35*(2), 243–268. https://doi.org/10.2308/ISYS-2020-017

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, *14*(1), 115–122. https://doi.org/10.2308/jeta-51730

Kouemou, G. L. (2011). History and theoretical basics of hidden markov models. In P. Dymarski (Ed.), *Hidden markov models: Theory and applications* (pp. 3–26) InTechOpen Access Publishers.

Krieger, F., Drews, P., & Velte, P. (2021). Explaining the (non-) adoption of advanced data analytics in auditing: A process theory. *International Journal of Accounting Information Systems*, *41*, Article 100511. https://doi.org/10.1016/j.accinf.2021.100511

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, 2, 396–404.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, *60*(3), 293–303. https://doi.org/10.1016/j.bushor.2017.01.004

Li, Q., & Liu, J. (2020). Development of an intelligent NLP-based audit plan knowledge discovery system. *Journal of Emerging Technologies in Accounting*, *17*(1), 89–97. https://doi.org/10.2308/jeta-52665

Liu, M., Robin, A., Wu, K., & Xu, J. (2021). Blockchain's impact on accounting and auditing: A use case on supply chain traceability. *Journal of Emerging Technologies in Accounting*, *19*(2), 105–119. https://doi.org/10.2308/JETA-2021-002

Liu, W.-P., Yen, M.-F., & Wu, T.-Y. (2022). Report users' perceived sentiments of key audit matters and firm performance: Evidence from a deep learning-based natural language processing approach. *Journal of Information Systems*, *36*(3), 191–209. https://doi.org/10.2308/ISYS-2020-061

Liu, X., Ghazali, K. H., Han, F., & Mohamed, I. I. (2021). Automatic detection of oil palm tree from UAV images based on the deep learning method. *Applied Artificial Intelligence*, *35*(1), 13–24. https://doi.org/10.1080/08839514.2020.1831226

Liu, Z., Huang, D., Huang, K., Li, Z. & Zhao, J. (2020). FinBERT: A pre-trained financial language representation model for financial text mining. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, 4513–4519. https://doi.org/10.24963/ijcai.2020/622

Lombardi, D. R., & Dull, R. B. (2016). The development of AudEx: An audit data assessment system. *Journal of Emerging Technologies in Accounting*, *13*(1), 37–52. https://doi.org/10.2308/jeta-51445

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Marten, K.-U., Quick, R., & Ruhnke, K. (2020). *Wirtschaftsprüfung: Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen* (6th ed.). Schäffer Poeschel.

McCallig, J., Robb, A., & Rohde, F. (2019). Establishing the representational faithfulness of financial accounting information using multiparty security, network analysis and a blockchain. *International Journal of Accounting Information Systems*, *33*, 47–58. https://doi.org/10.1016/j.accinf.2019.03.004

McCarthy, W. E. (2012). Accounting craftspeople versus accounting seers: Exploring the relevance and innovation gaps in academic accounting research. *Accounting Horizons*, *26*(4), 833–843. https://doi.org/10.2308/acch-10313

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Montazemi, A. R., & Qahri-Saremi, H. (2015). Factors affecting adoption of online banking: A meta-analytic structural equation modeling study. *Information & Management*, *52*, 210–226. https://doi.org/10.1016/j.im.2014.11.002

Moon, J. S., & Wood, D. A. (2020). Research initiatives in accounting education: Research relevance and research productivity. *Issues in Accounting Education*, *35*(4), 111–124. https://doi.org/10.2308/ISSUES-2020-012

Mousa, G. A., Elamir, E. A. H., & Hussainey, K. (2022). Using machine learning methods to predict financial performance: Does disclosure tone matter? *International Journal of Disclosure and Governance*, *19*, 93–112. https://doi.org/10.1057/s41310-021-00129-x

Nasir, M., Simsek, S., Cornelsen, E., Ragothaman, S., & Dag, A. (2021). Developing a decision support system to detect material weaknesses in internal control. *Decision Support Systems*, *151*, Article 113631. https://doi.org/10.1016/j.dss.2021.113631

Nehmer, R. A., & Srivastava, R. P. (2016). Using belief functions in software agents to test the strength of application controls: A conceptual framework. *International Journal of Intelligent Information Technologies*, *12*(3), 1–19. https://doi.org/10.4018/IJIIT.2016070101

No, W. G., Lee, K. K., Huang, F., & Li, Q. (2019). Multidimensional audit data selection (MADS): A framework for using data analytics in the audit data selection process. *Accounting Horizons*, *33*(3), 127–140. https://doi.org/10.2308/acch-52453

Nonnenmacher, J., Kruse, F., Schumann, G., & Gómez, J. M. (2021). Using autoencoders for data-driven analysis in internal auditing. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 5748–5757. http://hdl.handle.net/10125/71317

O'Leary, D. E. (2019). Some issues in blockchain for accounting and the supply chain, with an application of distributed databases to virtual organizations. *Intelligent Systems in Accounting, Finance and Management*, *26*(3), 137–149. https://doi.org/10.1002/isaf.1457

Otero, A. R. (2015). An information security control assessment methodology for organizations' financial information. *International Journal of Accounting Information Systems*, *18*, 26–45. https://doi.org/10.1016/j.accinf.2015.06.001

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*, 183–199. https://doi.org/10.1016/j.im.2014.08.008

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129–139. https://doi.org/10.1080/0960085X.2018.1458066

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Perols, J. L., & Murthy, U. S. (2012). Information fusion in continuous assurance. *Journal of Information Systems*, *26*(2), 35–52. https://doi.org/10.2308/isys-50216

Plant, O. H., van Hillegersberg, J., & Aldea, A. (2022). Rethinking IT governance: Designing a framework for mitigating risk and fostering internal control in a DevOps environment. *International Journal of Accounting Information Systems*, *45*, Article 100560. https://doi.org/10.1016/j.accinf.2022.100560

Polizzi, S., & Scannella, E. (2023). Continuous auditing in public sector and central banks: A framework to tackle implementation challenges. *Journal of Financial Regulation and Compliance*, *31*(1), 40–59. https://doi.org/10.1108/JFRC-02-2022-0011

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106. https://doi.org/10.1007/BF00116251

Rabiner, L., & Juang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, *3*(1), 4–16. https://doi.org/10.1109/MASSP.1986.1165342

Rahimian, F., Bajaj, A., & Bradley, W. (2016). Estimation of deficiency risk and prioritization of information security controls: A data-centrix approach. *International Journal of Accounting Information Systems*, *20*, 38–64. https://doi.org/10.1016/j.accinf.2016.01.004

Rajgopal, S. (2021). Integrating practice into accounting research. *Management Science*, *67*(9), 5430–5454. https://doi.org/10.1287/mnsc.2020.3590

Ralyté, J., Deneckère, R., & Rolland, C. (2003). Towards a generic model for situational method engineering. In J. Eder & M. Missikoff (Eds.), *Advanced information systems engineering. CAiSE 2003. Lecture notes in computer science* (Vol. 2681, pp. 95–110). Springer. https://doi.org/10.1007/3-540-45017-3_9

Raschke, R. L., Saiewitz, A., Kachroo, P., & Lennard, J. B. (2018). AI-enhanced audit inquiry: A research note. *Journal of Emerging Technologies in Accounting*, *15*(2), 111–116. https://doi.org/10.2308/jeta-52310

Rish, I. (2001). An empirical study of the naive bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 41–46. https://sites.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf

Ritchi, H., Bandana, A., Zaldy, A., & Alfian, A. (2022). Permissioned blockchain for business process visibility: A case of expenditure cycle. *Procedia Computer Science*, *197*, 336–343. https://doi.org/10.1016/j.procs.2021.12.148

Rozario, A. M., & Thomas, C. (2019). Reengineering the audit with blockchain and smart contracts. *Journal of Emerging Technologies in Accounting*, *16*(1), 21–35. https://doi.org/10.2308/jeta-52432

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1987). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Foundations* (Vol. 2, pp. 318–362). The MIT Press.

Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education Limited.

Saeedi, A. (2021). Audit opinion prediction: A comparison of data mining techniques. *Journal of Emerging Technologies in Accounting*, *18*(2), 125–147. https://doi.org/10.2308/JETA-19-10-02-40

Salijeni, G., Samsonova-Taddei, A., & Turley, S. (2021). Understanding how big data technologies reconfigure the nature and organization of financial statement audits: A sociomaterial analysis, *European Accounting Review*, *30*(3), 531–555. https://doi.org/10.1080/09638180.2021.1882320

Schapire, R. E., & Singer, Y. (1999). Improved boosting algorithms using confidence-rated predicitions. *Machine Learning*, *37*(3), 297–336. https://doi.org/10.1023/A:1007614523901

Schreyer, M., Sattarov, T., Borth, D., Dengel, A., & Reimer, B. (2018). Detection of anomalies in large-scale accounting data using deep autoencoder networks. *arXiv preprint*. https://doi.org/10.48550/arXiv.1709.05254

Schreyer, M., Sattarov, T., Gierbl, A., Reimer, B., & Borth, D. (2020). Learning sampling in financial statement audits using vector quantised autoencoder neural networks. *Proceedings of the First ACM International Conference on AI in Finance (ICAIF'20)*, 1–8. https://doi.org/10.1145/3383455.3422546

Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21. https://doi.org/10.1145/3068335

Schultz, M., & Tropmann-Frick, M. (2020). Autoencoder neural networks versus external auditors: Detecting unusual journal entries in financial statement audits. *Proceedings of the 53rd Hawaii International Conference on Systems Sciences*, 5421–5430. https://hdl.handle.net/10125/64408

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A hybrid approach to alleviating class Imbalance. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, *40*(1), 185–197. https://doi.org/10.1109/TSMCA.2009.2029559

Sheldon, M. D. (2021). Auditing the blockchain oracle problem. *Journal of Information Systems*, *35*(1), 121–133. https://doi.org/10.2308/ISYS-19-049

Siano, F., & Wysocki, P. (2021). Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small(er) data sets. *Accounting Horizons*, *35*(3), 217–244. https://doi.org/10.2308/HORIZONS-19-161

Sifa, R., Ladi, A., Pielka, M., Ramamurthy, R., Hillebrand, L., Kirsch, B., Biesner, D., Stenzel, R., Bell, T., Lübbering, M., Nütten, U., Bauckhage, C., Warning, U., Fürst, B., Khameneh, T. D., Thom, D., Huseynov, I., Kahlert, R., Schlums, J., ... Loitz, R. (2019). Towards automated auditing with machine learning. *Proceedings of the ACM Symposium on Document Engineering 2019 (DocEng'19)*, 1–4. https://doi.org/10.1145/3342558.3345421

Singh, V. K., & Dutta, K. (2018). Predicting security events in cloud computing: Completed research. Twenty-fourth Americas Conference on Information Systems, New Orleans, LA, USA.

Sonnenberg, C., & vom Brocke, J. (2012). Evaluations in the science of the artificial – Reconsidering the build-evaluate pattern in design science research. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design science research in information systems. Advances in theory and practice. DESRIST 2012. Lecture notes in computer science* (Vol. 7286, pp. 381–397). Springer. https://doi.org/10.1007/978-3-642-29863-9_28

Summers, S. L., & Wood, D. A. (2017). An evaluation of the general versus specialist nature of top accounting journals. *Accounting Horizons*, *31*(2), 105–124. https://doi.org/10.2308/acch-51712

Sun, T. (2018). The incremental informativeness of the sentiment of conference calls for internal control material weaknesses. *Journal of Emerging Technologies in Accounting*, *15*(1), 11–27. https://doi.org/10.2308/jeta-51969

Sun, T. S. (2019). Applying deep learning to audit procedures: An illustrative framework. *Accounting Horizons*, *33*(3), 89–109. https://doi.org/10.2308/acch-52455

Szeliski, R. (2022). *Computer vision: Algorithms and applications* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-34372-9

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson Education Limited.

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., ... Le, Q. (2022). LaMDA: Language models for dialog applications. *arXiv preprint*. https://doi.org/10.48550/arXiv.2201.08239

Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing and Management*, *43*(5), 1216–1247. https://doi.org/10.1016/j.ipm.2006.11.011

van der Aalst, W. (2016). *Process mining: Data science in action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4_1

van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128–1142. https://doi.org/10.1109/TKDE.2004.47

van der Aalst, W. M. P. (2010). Process discovery: Capturing the invisible. *Institute of Electrical and Electronics Engineers (IEEE) Computational Intelligence Magazine*, *5*(1), 28–41. https://doi.org/10.1109/MCI.2009.935307

Vanwersch, R. J. B., Shahzad, K., Vanderfeesten, I., Vanhaecht, K., Grefen, P., Pintelon, L., Mendling, J., van Merode, G. G., & Reijers, H. A. (2016). A critical evaluation and framework of business process improvement methods. *Business and Information Systems Engineering*, *58*(1), 43–53. https://doi.org/10.1007/s12599-015-0417-x

Vapnik, V. N. (2000). *The nature of statistical learning theory* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4757-3264-1

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons*, *29*(2), 381–396. https://doi.org/10.2308/acch-51071

vom Brocke, J., Baier, M.-S., Schmiedel, T., Stelzl, K., Röglinger, M., & Wehking, C. (2021). Context-aware business process management: Method assessment and selection. *Business and Information Systems Engineering*, *63*(5), 533–550. https://doi.org/10.1007/s12599-021-00685-0 vom Brocke, J., Riemer, K., Plattfaut, R., Simons, A., Niehaves, B., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems 37*(9), 205–224. https://doi.org/10.17705/1CAIS. 03709

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: on the importance of rigour in documenting the literature search process. *Proceedings of the 17th European Conference on Information Systems (ECIS 2009), 161*. https://aisel.aisnet.org/ecis2009/161

Wang, Y., Chiu, T., & Chiu, V. (2020). Redesigning business process to comply with the wew revenue recognition standard using process mining. *Journal of Emerging Technologies in Accounting*, *17*(1), 149–163. https://doi.org/10.2308/jeta-52663

Warren, J. D., Moffitt, K. C., & Byrnes, P. (2015). How big data will change accounting. *Accounting Horizons*, *29*(2), 397–407. https://doi.org/10.2308/acch-51069

Waymire, G. B. (2012). Introduction for essays on the state of accounting scholarship. *Accounting Horizons*, *26*(4), 817–819. https://doi.org/10.2308/acch-50236

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Werner, M. (2017). Financial process mining – Accounting data structure dependent control flow inference. *International Journal of Accounting Information Systems*, *25*, 57–80. https://doi.org/10.1016/j.accinf.2017.03.004

Werner, M. (2019). Materiality maps – Process mining data visualization for financial audits. *Proceedings of the 52nd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences*, 1045–1054. https://doi.org/10125/59544

Werner, M., & Gehrke, N. (2015). Multilevel process mining for financial audits. *IEEE Transactions on Services and Computing*, *8*(6), 820–832. https://doi.org/10.1109/TSC.2015.2457907

Werner, M., & Gehrke, N. (2019). Identifying the absence of effective internal controls: An alternative approach for internal control audits. *Journal of Information Systems*, *33*(2), 205–222. https://doi.org/10.2308/isys-52112

Werner, M., Wiese, M., & Maas, A. (2021). Embedding process mining into financial statement audits. *International Journal of Accounting Information Systems*, *41*, Article 100514. https://doi.org/10.1016/j.accinf.2021.100514

Williams, G., Baxter, R., He, H., Hawkins, S., & Gu, L. (2002). A comparative study of RNN for outlier detection in data mining. *Proceedings of the 2002 IEEE International Conference on Data Mining*, 709–712. https://doi.org/10.1109/ICDM.2002.1184035

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45–55. https://doi.org/10.1057/ejis.2011.51

Wood, D. A. (2016). Comparing the publication process in accounting, economics, finance, management, marketing, psychology, and the natural sciences. *Accounting Horizons*, *30*(3), 341–361. https://doi.org/10.2308/acch-51443

World Economic Forum (2015). *Deep shift: Technology tipping points and societal impact*. Retrieved March 16, 2024 from https://www.weforum.org/publications/deep-shift-technology-tipping-points-and-societal-impact/

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

Yan, J., Wang, X., Wang, B., & Zhang, Y. (2019). Research on application of data mining technology in risk assessment process of audit. *2019 International Conference on Economic Management and Model Engineering (ICEMME)*, 487–491. https://doi.org/10.1109/ICEMME49371.2019.00102

Zhang, C. A., Thomas, C., & Vasarhelyi, M. A. (2022). Attended process automation in audit: A framework and a demonstration. *Journal of Information Systems*, *36*(2), 101–124. https://doi.org/10.2308/ISYS-2020-073

Zhang, G., Atasoy, H., & Vasarhelyi, M. A. (2022). Continuous monitoring with machine learning and interactive data visualization: An application to a healthcare payroll process. *International Journal of Accounting Information Systems*, *46*, Article 100570. https://doi.org/10.1016/j.accinf.2022.100570

Zhaokai, Y., & Moffitt, K. C. (2019). Contract analytics in auditing. *Accounting Horizons*, *33*(3), 111–126. https://doi.org/10.2308/acch-52457

Tassilo Lars Föhr is a Doctoral Candidate at the Institute of Accounting and Auditing at the University of Ulm, Germany.

Address: Institute of Accounting and Auditing, Helmholtzstrasse 22, 89081 Ulm, Germany, Tel: +49 731 50 33001, E-mail: tassilo.foehr@uni-ulm.de

Acknowledgements: I am very grateful for comments received from Kai-Uwe Marten, Marco Schreyer, Marc Eulerich, Carla Wilkin, David Wang, and the participants of the 2023 Joint Midyear Meeting of the AIS, SET and International Sections in Las Vegas, NV, USA.