www.cya.unam.mx/index.php/cya

Contaduría y Administración
Accounting & Management

Contaduría y Administración 70 (2), 2025, 184-219

Big data analytics for financial auditing practices: Identification of conceptual patterns, implications and challenges using text mining

Análisis de big data para prácticas de auditoría financiera: identificación de patrones conceptuales, implicaciones y desafíos mediante la minería de textos

Kamakshaiah Musunuru *

GITAM School of Business, India

Received November 6, 2023; accepted April 24, 2024
Available online May 31, 2024

Abstract

Big data analytics and the practice of related technologies is rampant in the corporate world across the globe. The ability of companies to collect, store and analyze massive amounts of data and use such data for decisions is considered critical for a firm's success. The auditing industry is not up to the mark, lacks sufficient emphasis and practice of big data analytics. This study assumes that big data processing technologies can impact financial auditing practices positively. Data sets were mined from literature using text mining methodology. Conceptual patterns such as Auditing, Fraud, Risk, and Security were found to be highly influential in the literature. Opinion in the literature is diverge for conceptual patterns such as Auditing, Fraud, and Risk but not for Security. Few potential implications under four main categories such as technologies, enablers, challenges, and compliance were identified. Digital technologies, specifically Artificial Intelligence (AI) and Blockchains were found to be enablers for firm's performance and growth. Fraud detection, forensics, legitimacy were found to be few challenges for compliance. In addition to this, big data analytics (BDA) was found to be a moderating variable for technologies like Blockchains, Artificial Intelligence (AI) and for challenges such as Risk, Security but not for Fraud.

JEL Code: M4, O3
Keywords: big data analytics; financial auditing; text mining; literature review; data driven decision making

*Corresponding author.
E-mail address: kmusunur@gitam.edu (K. Musunuru).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

http://dx.doi.org/10.22201/fca.24488410e.2025.5283

Resumen

El análisis de big data y la práctica de tecnologías relacionadas están muy extendidos en el mundo empresarial de todo el mundo. La capacidad de las empresas para recopilar, almacenar y analizar cantidades masivas de datos y utilizarlos para tomar decisiones se considera fundamental para el éxito de una empresa. La industria de la auditoría no está a la altura y carece de suficiente énfasis y práctica en el análisis de big data. Este estudio supone que las tecnologías de procesamiento de big data pueden impactar positivamente las prácticas de auditoría financiera. Los conjuntos de datos se extrajeron de la literatura utilizando la metodología de minería de textos. Se encontró que patrones conceptuales como Auditoría, Fraude, Riesgo y Seguridad eran muy influyentes en la literatura. Las opiniones en la literatura son divergentes respecto de patrones conceptuales como Auditoría, Fraude y Riesgo, pero no respecto de la Seguridad. Se identificaron pocas implicaciones potenciales en cuatro categorías principales, como tecnologías, facilitadores, desafíos y cumplimiento. Se descubrió que las tecnologías digitales, específicamente la Inteligencia Artificial (IA) y las cadenas de bloques, facilitaban el desempeño y el crecimiento de las empresas. Se descubrió que la detección de fraude, la ciencia forense y la legitimidad eran pocos desafíos para el cumplimiento. Además de esto, se descubrió que el análisis de big data (BDA) es una variable moderadora para tecnologías como Blockchains, Inteligencia Artificial (IA) y para desafíos como Riesgo y Seguridad, pero no para el Fraude.

Código JEL: M4, O3
Palabras clave: análisis de grandes datos; auditoría financiera; extracción de textos; revisión de literatura; toma de decisiones basada en datos

Introduction

Big data is the new wave for data strategy and operational performance in the industry (Nguyen et al, 2024; Jin et al., 2015). Customer demands are evolving and making better decisions by leveraging the volume, variety and velocity of big data is critical for effective operational costs and financial performance (Ochuba et al., 2024; Johnson et al., 2017). Today big data analytics and practice of related technologies are rampant across industries (Ülkü and Mansouri, 2024; McNeely and Hahm, 2014). The ability of the companies to collect data massively and analyze such data in real-time scenario is considered a critical strength. Data drives growth and is considered on par with dollars in the corporate world (Kazemi et al, 2024; Khan et al., 2014). Today data has grown massively and utilization of the same across every echelon of the organization is significant (Schildt, 2017). Predictive and prescriptive analytics are used to estimate the likelihood of future outcomes based on historical data. Artificial Intelligence (AI) and Machine Learning (ML) coupled with big data analytics has accelerated such analytics in the business industry (Amir et al., 2022). Big data as a collection of technologies might be useful for event predictions and preventive measures (Anitha et al., 2024; Appelbaum et al., 2017). There is enormous evidence in the literature in support of the symbiotic relationship between artificial intelligence and big data and their impact on auditing practice. For instance, a few auditing challenges associated with case detection and

medical treatment during Covid-19 were addressed using artificial intelligence and big data technologies (Alharasis et al., 2024). Few similar studies were provided in the forthcoming sections.

Unfortunately, the financial auditing industry is not up to the trends, lacks practices of big data analytics and associated technologies (Azzam et al., 2024). Partly this might be due to auditors' oversight of information technology. The point is not about relevance or applicability but the ability of the auditors in leveraging data centric decision making using real-time data and processing the same for compliance and value. Big data analytics is the branch of data analytics as such it is not extraneous to admit the fact that data driven decision making is not a foreign idea, and then big data processing technology is need of the hour for financial auditing practitioners (Ahmadi, 2024; Ahmad, 2024). Collecting, storing, and analyzing data are few organizational competencies and key success factors. If the main purpose of auditing is to ensure compliance and accountability then big data technologies as being a close cousin of data processing can ensure increased fairness and trust in the business (Olabanji et al., 2024). Furthermore, auditors and other stakeholders will have a higher level of confidence when it comes to analyzing the effectiveness of specific company operations (Smith, 2018). The role and applications of big data in auditing can be challenging (Hima, 2024; Dagilienė and Klovienė, 2019). However, by accommodating big data analytics as in organization strategy may make it possible to unleash the power of data processing and ensuring data driven decision making in real time setups (Hashem et al., 2016).

With this backdrop and given the present state of the auditing industry, this paper tries to find the implications of big data technologies and their impact on auditing practices. Apart from this, this paper also seeks to find a few associated technologies and their relationship with big data analytics (BDA) in the literature.

Review of literature

The term 'Big Data' along with other trending topics such as 'Data Science' and 'Artificial Intelligence (AI)' have become buzzwords in the accounting profession in recent years (Abu Huson et al., 2024). Data analytics is a relatively a new skillset which is growing significantly in this area of accounting and auditing industry (Rijanto, 2024; Alzahrane, 2024). Therefore, accounting professionals can capitalize on numerous opportunities in this rapidly evolving, disruptive, but ultimately advantageous environment by embracing big data, data analytics, and artificial intelligence (AI) to stay ahead in the industry (Rana et al., 2023). There is lot of evidence in the literature in support of aforementioned statements. For instance, Sun et al. (2024) argue that the accounting and auditing fields can utilize big data processing technologies to ensure accuracy, transparency, and integrity in financial reporting field. The authors indeed suggested few other emerging technologies such as Artificial Intelligence (AI), Machine Learning (ML), Internet of

Things (IoT) together with big data for effective financial auditing practices. Sanoran and Ruangprapun (2023) did certain qualitative study on twenty-eight audit professionals to gain insights into their experience implementing audit data analytic tools. Authors finds that the auditors perceive about benefits of audit data analytic tools in improving their audit process management. Bose et al. (2023) identifies the growing role played by big data, data analytics, and artificial intelligence (AI) in the accounting profession. There are number of studies which argues that new technologies will have a positive impact on auditors as they will be allowed to focus on value-added activities (Vitali and Giuliani, 2024; Abu Huson et al., 2024; Wassie and Lakatos, 2024; Rahman and Irwansyah, 2024; Oluwagbade et al., 2024)

Malsch and O'Dwyer (2021) studied the impact of big data technologies on the organization of financial audits. The study shows how visualization dashboards have contributed to auditors' ability to communicate and justify their claims and judgements. Authors demonstrated the ability of BDA in reshaping the nature of work relationships and flows between audit firms' different functions and service lines. George et al. (2016) studied big data changes in audit technology focusing on three key aspects, namely the impact of BDA on the nature of the relationship between auditors and their clients. The study identified consequences of the technology for the conduct of audit engagements besides common challenges associated with embedding BDA in the audit context.

From the above review it appeared that "Fraud", "Risk" and "Security" were key issues for financial auditors. It is also clear that BDA is important for financial auditing for detecting fraud, mitigating risk and as a security tool. However, the level of influences or impacts between BDA and these issues are not clear in the current research. This study admits these issues as challenges and tries to identify evidence in the literature using text mining and analysis.

Big data technologies for auditing practices

Implementation of information systems is still a nascent idea in auditing industry (Rijanto, 2024). Auditing involves huge amount of data processing and impact of information retrieval, storage, analysis, and reporting are critical for organizational success (Ionescu et al, 2014). If data processing is critical for auditing practices, then big data technology ensures efficiency in such processing. In fact, Sihombing et al. (2023) through their experimental study on 109 accounting participants finds that auditors' skills and knowledge of data analytics and big data can influence their planning skills and the auditors also will be able to determine the level of audit risk and estimate audit duration accurately. Patel and Shah (2023) performed regression analysis together with a moving average estimate to quantify the auditing industry's dependence on bid data technology and measured the impact of big data on auditing efficiency. With this

notion this paper tries to explore literature assuming a significant relationship between big data technology and its impact on auditing practices.

Technologies for streamlining auditing services

Technology practices and their impact on management accountants and accounting is rising continuously. Hezam et al. (2024) reviewed a total of 100 articles from Web of Science and Scopus. The analysis reveals that data analytics is a promising domain for auditing practice as it improves audit efficiency and promotes audit work digital transformation. This study offers insights on new directions and views on big data analytics in auditing.

Sangster et al. (2009) studied the impact of ERP on accounting practices in large UK-based organizations. They found that under successful ERP implementations, management accountants have time for other, less mundane activities and their role becomes more enriching and the reverse is not true. This study appears to be a remarkable one pertaining to Accounting Information Systems (AIS). In another similar study, Grabski, et al. (2011) finds ERP as the most demanding information systems implemented for accounting practices, representing a major departure from the individual and departmental information systems prevalent in the past. This study brings enormous emphasis on technologies touting as a transformative force in the accounting profession. Almasria et al. (2021) investigated the role of AIS in enhancing the quality of external audit procedures. The investigation illustrated elements and applications of AIS and their contribution to the quality of the external audit procedures. The study concludes that the audit process, audit testing procedures, risk assessments, implementation are critical for high-quality auditing practices which involve information systems. Studies that were examined until here indicate the importance of information systems particularly for accounting and auditing practices. The phrase Accounting Information Systems (AIS) is so generic and might refer to any type of applications particularly related to Information Technology (IT). In addition, IT might refer to a plethora of applications such as artificial intelligence, machine learning, Blockchains, internet of things (IoT) and many more. The following subsections provide insights related to implications of these technologies.

Artificial intelligence

Artificial Intelligence (AI) and Machine Learning (ML) can address a few challenges of auditing practices. Oluwagbade et al. (2024) emphasizes the importance of AI and ML for auditors and business firms. The study finds that the impact of machine learning on accounting practices is positive and statistically significant. Kanaparthi (2024) observes that the AI can automate repetitive financial accounting tasks and

assists organizations in circumventing the need for additional staff, thereby minimizing associated costs. The author states that as a matter of consequence, businesses are increasingly embracing blockchain technology and AI applications in their financial accounting operation to bolster efficiency. Hasan (2021) performed a semi-systematic review on challenges of disruptive technologies brought forth by the industry 4.0 over auditing discipline. The review makes the affirmation that the wider application of AI in accounting and auditing profession provides the benefits of greater efficiency, productivity, and accuracy. Authors also warn about the burden of income and wealth inequality, extinction of traditional jobs and unskilled workforce.

Xie (2019) alluded as Bayesian model provides objective and rational probability to auditor and help them to make more accurate assessment, as well as reducing the misjudgments caused by auditor's personal emotion. The author further emphasizes the importance of the model quoting that it is a popular machine learning algorithm that has become so famous and has been widely used in financial area until now. Pothumsetty, (2020) studied the role of AI and its impact on replacing jobs of auditors, accountants, and finance professionals. The author identified and picked articles from various research databases such as Jstor, google scholar and proquest etc., and performed analysis required for his research. This paper focuses on the various benefits AI and their significance to finance companies and professionals and concludes that AI software will be able to handle most of the monotonous accounting and auditing tasks faster compared to human beings but there is always a need for human beings to interpret the data that is generated from AI technology.

By and large, AI appears to impact auditing practices. The impact is ascertained by algorithms, increased efficiency and productivity, replacement of human auditing skills and the auditing profession at large. The current study tries to find supporting evidence in the literature using text mining and analysis.

Blockchains

Auditing practices were identified as impacted by Blockchains and other technologies in the previous sections. This subsection finds the rationale in support of those learnings. Bejjar and Siala (2024) find Blockchains as beneficial for automated audit processes, improved risk management, cost savings, fraud reduction, besides ensuring better auditing practices in business firms. Bonyuet (2020) reviewed research on Blockchain technology and assessed the impact of Blockchain in the audit profession, including new risks, changes in procedures and additional opportunities. The author concludes that the auditors are business consultants and therefore, are expected to be subject matter experts in all relevant technologies. As a result, even if the impact of Blockchain on auditing is minimal, auditors need to understand how such technology impacts client's business. Liu et al. (2022) observes that Blockchain technology has the

potential to change the landscape of the accounting profession in general and auditing services in particular. The authors recommended auditors to adapt, adjust, and elevate themselves to the role of strategic partners in Blockchain implementation. Brender et al. (2024) argues that the characteristics of the Blockchain technology viz., transparency, traceability and integration of rules and procedures in the technology, combined with current technical developments, process and service innovation, applications such as smart contracts and publicly held registers, have the potential to significantly change audit and control activities. Anwar et al. (2019) critically analyzed the impacts and scope of Blockchain technology for auditing with the strong literature review in the field of Blockchain technology and its application in the audit process. The study provided a framework for auditing financial transactions using Blockchain technology. These authors concluded that Blockchain technology provides a very secure platform for the transaction and at the very same time auditing can also be completed due to transparency of Blockchain technology. Desplebin et al. (2021) supports Blockchain technology affirming its innovative potential in the fields of accounting and auditing. The study identified three major challenges regarding the future of Blockchain in accounting and auditing and they are (i) the transformation of accounting techniques; (ii) main evolutions in accounting and auditing; and (iii) main evolutions in the work, skills, and education of auditors. By and large there is potential evidence in the literature in support of Blockchain technologies.

Most of the studies argue in support of Blockchains and their implementation for long run, secure, transparent, error free auditing practices while highlighting challenges such as transformation of legacy systems, hiccups in evolution, educating auditors and development of technical skills.

The benefits of using big data in financial auditing

The advantages of big data in auditing are aplenty. However, many companies have not figured out how to fully leverage the power of big data to make auditing more accurate, more reliable, and less costly (Rahman and Irwansyah, 2024). Collecting, aggregating, and analyzing data in a manner that provides valuable insights are critical to financial auditors (Hosseinpour and Jans, 2024). However, that need to devote adequate resources to support internal auditing tasks. Providing access to critical data and collaborating with multiple departments will help auditors prepare more accurate, dependable, and timely reports. Big data technologies can ensure seamless network of databases that were otherwise local to each of the departments. Abdelwahed et al. (2024) present a few findings regarding the use of Big Data (BD) and Big Data Analytics (BDA) in auditing. The findings disclose a comprehensive view of current practices by studying practices at a few auditing firms. The study finds that BD and BDA could impact auditing practices when data analysis is carried out on non-financial data and new data items as complementary to financial data. A few studies reveal significant positive associations between BD

practices and firm performance, earnings determinants, reporting quality, fair value measurements, firm value, risk management, the efficiency of the decision-making process, narrative disclosure, and firm sustainability (Azzam et al., 2024; Peters and Onamariwari, 2024). The main goal of the current study is to find the consequences of big data and accounting standards. The literature supports change in standards to focus on data, the processes that generate them, and their analysis, rather than their presentation, will add value and relevance to the accounting profession.

Big data involves the use of automation and artificial intelligence, data can be processed in larger volumes and higher velocity to uncover valuable insights for auditors (Chu and Yong, 2021). For example, non-compliance, current policy changes, and fraud can be identified and used to guide the focus of both internal and external auditors. Fraud detection is one of the critical tasks in both financial and non-financial organizations (Jan, 2018). Big data can help financial auditors to streamline the reporting process and detect fraud. Auditing professionals can identify business risks in time and conduct more relevant and accurate audits (Manita et al., 2020). This requires adequate data management and aggregation processes in place, and it needs to be part of the financial strategy of the firm. Data sets that were collected and analyzed for the purpose of auditing need to be checked for accuracy, timeliness, and capacity. In this way, decisions that auditors make are based on reliable and high-quality information. This is particularly important when it comes to auditing because decisions regarding compliance, risk, and investment are typically made after audit reports are prepared (Knechel and Salterio, 2016). Big data also allows auditors to automate multiple portions of the auditing process (Appelbaum et al., 2017). Human error is one of the reasons that affects compliance. By automating manual and repetitive tasks, auditors can set up various controls in advance and monitor how well a company is adhering to established guidelines. Ioannou et al, (2021) are of the opinion that the audit's purpose is to examine the true and fair view of financial statements and is increasing in complexity. The authors notes that this complexity is arising due to deluge of financial data in manifolds. Exactly, at this conjuncture, the demand for technologies which can handle real time, yet massive amounts of data will arise.

There are another side of arguments either. The view that big data analytics are applicable to auditing process is still an unacceptable idea among stakeholders. De Santis and D'Onza (2021) studied BDA's legitimation processes in the audit professional environment. The authors found that BDA complements traditional audit procedures inside the organization. However, the digital maturity of audit clients, the lack of audit standards and the audit oversight authority's negative view prevent the legitimation of BDA. Alrashidi et al. (2022) studied the impact of BDA on continuous auditing. The authors point out four essential gaps viz., data consistency, integrity, aggregation, identification, and confidentiality which makes BDA feasible for auditing. Alles and Gray (2016) point out that application of BDA to auditing is less and alludes upon one of the BDA characteristics known as "veracity" for such

failure. The study complains that veracity relates to the integrity of the data and remains problematic to auditors because developing appropriate level of confidence in auditors is a big question.

Research methods

This study is basically descriptive in nature. The purpose of this study is to find the importance of big data analytics for financial auditing practices. The importance of big data analytics is ascertained by mining existing literature for a few conceptual patterns, challenges or issues and implications. Descriptive analysis was deemed to be fit for such exercise. Multiple regression was used for evaluating impact between big data analytics and financial auditing. This study also assumes certain moderating effects explained by big data analytics while testing through multiple regression analysis. Data was collected from Scopus as in abstracts. These abstracts were analyzed using statistical techniques to realize the study objectives. At first the abstracts were converted to term vectors and document term matrices through text mining methodologies, such data vectors and matrices were analyzed using text analytics subsequently (Musunuru, 2021).

The existing literature enlightens a few challenges that interplay between big data technologies and financial auditing. These challenges largely deal with certain critical aspects of financial auditing such as financial fraud, financial risk, and data security etc. based on the learnings from the literature, the following questions were brought forward for further investigation.

1. How big data technologies are implied in the financial auditing industry? What are those implications?
2. Are there any relevant or supportive technologies which are associated with big data analytics? What are the implications of those technologies?
3. How do big data technologies address challenges such as “fraud”, “risk”, “security”, in the auditing industry?
4. How do firms strategize performance and drivers for growth?

The purpose of financial auditing is to ensure the firm’s compliance and commitment for fairness and communicate the same to stakeholders to create trust in the way business is being operated. The main aim of this study is to seek the influence or impact of big data technologies on financial auditing services. Hence, the study assumes that big data analytics and related technologies can ensure effective and efficient financial auditing practices.

Sampling procedure and data sets

The above research questions provide inputs for data retrieval strategy. Data required for analysis were retrieved using a search string through Boolean operators. Usually, there are three main commands used to do this: AND, OR and NOT. These are referred to as Boolean operators. Data bases interpret these operators as a command instead of search terms when used in search query. "AND" is used to combine *different* topic words so that *all* the words must appear in the search results and the result will be intersection of sources; whereas, "OR" is used to combine topic words so that all the words can appear in the search results where the results will be union of sources (UoB, n.d.). Most databases do not allow "NOT" to exclude a term that doesn't appear (Atkinson and Cipriani, 2018; Bramer et al., 2018).

A common yet generic search statement or query was constructed in the light of study's aim and the learnings from the review of literature. The statement - "big data applications for financial auditing practices" - appeared as a plausible statement for further investigation. The search query was executed in Scopus database and the query retrieved 32 different yet potential documents. [1] Scopus allows only AND rules and not the other two i.e., OR and NOT. As a result, the data retrieved through literature search entails all those sources which have one of few search words of the statement or query. So that the data mining is done on all those sources which are the whole population instead of a sample. The data retrieval activity was done between October and November 2023. The data associated with these documents were arranged as in data matrix of the order 32 X 18, where all the rows were documents, and the columns were attributes associated with each document viz.,

Authors, Title, Year, Source title, Volume, Issue Art. No., Page start, Page end, Page count, Link, Abstract, Author Keywords, Index, Keywords, Publisher, Document Type, Publication Stage, Source.

The column named "Abstract" is the most critical attribute among the rest in the data matrix. The text available through abstracts across the sampled documents serves as text corpus. The analyses were performed using certain descriptive statistical methods coupled with few statistical tests and results were interpreted in the "Data Analysis" section.

[1] Data sets were kept available at https://github.com/Kamakshaiah/auditing-data-analysis which is companion repository for this article. List of abstracts and other information is available in the data file located at https://github.com/Kamakshaiah/auditing-data-analysis/blob/main/datasets/main-datasets/big-data-fin-auditing.csv in the repository.

Text mining & analysis

The text corpus was created from the list of abstracts from sampled documents. Text mining and analysis was performed, on the text corpus, using LitReviewer module (Musunuru, 2021).[2] LitReviewer is a Python based module which is used for analyzing data sets created from text corpuses. LitReviewer helps in performing data mining and analysis using methods which belong to the area of Natural Language Processing (NLP). Text mining is the process of deriving high-quality information from the text corpus. In theory, text mining involves "the discovery of previously unknown information, by automatically extracting the same from different written resources."[3] Written resources may include a collection or corpus of text as in documents retrieved through several sources such as websites, books, emails, reviews, and articles. High-quality information is typically obtained by devising patterns and trends such as statistical pattern learning.

In this study, the word “pattern” refers to the word associated with study issue or research question. These challenges or questions were studied through the word conjugates. In fact, the word “morpheme” is more apt than the word “conjugate” for the present context of the study. [4] However, the word “conjugate” is used to keep the study less vigorous while dealing with technicalities. In text mining, the word “conjugate” refers to a word or term which can be coupled, connected, or related to another word or term. Text analytics usually involves the process of structuring the input text, deriving patterns within the structured data, leading to evaluation and interpretation of the output. While text mining can disinter invisible patterns in the data, but text analysis is useful for data analysis on such patterns.

Sparse data sets

Data which is extracted from text mining is usually sparse in nature. This study uses sparse data sets which were obtained from text mining of text corpus. Sparse data is any data which does not contain actual data related to the phenomenon of the study and possess mostly empty or have zero value across the data

[2] Available at https://github.com/Kamakshaiah/literature-review.

[3] Hearst, M. (2003). What is text mining. SIMS, UC Berkeley, 5. Available at https://www.jaist.ac.jp/~bao/MOT-Ishikawa/FurtherReadingNo1.pdf.

[4] Verb conjugation exhibits rich morphological variations in text corpus. For instance, Marquis, & Shi, (2009) studied recognition of verb roots and bound morphemes. The authors observes that morphological variations like allophonic alternations and syllabification arises out of verb conjugations. Hence, conjugates and morphemes can be two complementary aspects in the context of text mining.

variables. As such, sparse data is different from missing data because sparse data shows up as empty or zero while missing data doesn't show values and any such absence of data is denoted or represented by NAs and NANs. Sparse data sets are usually handled through outlier analysis coupled with box plots to trace meaningful patterns (Qu and Genton, 2022). Outlier analysis detects actual instances or cases of data and box plots are useful to filter such instances further by increasing the relative value of the same in comparison to other instances of the samples. In this study, documents which were collected from Scopus, through appropriate search statements, are used for creating text corpus. This text corpus was mined for "conceptual patterns" using LitReviewer package. As a result, a few data sets were obtained which are essentially sparse in nature. These sparse data sets were analyzed using a few statistical tests related to normality, and heterogeneity. The master data file together with sparse data sets are available from companion repository https://github.com/Kamakshaiah/auditing-data-sets for this article.

Document Term Matrices (DTM)

Document Term Matrices (DTMs) were prepared using Python's LitReviewer module. The "textmining" script in this module has methods to prepare DTM for any given word pattern.[5] For instance, the audit data file is a set of term vectors for various conjugates of term "audit". [6]

Table 1
Data files with word patterns with respective conjugates

Data file [a]	Dimensionality [b]	Word conjugates [c]
Audit	32 X 9	Audit, auditability, auditable, audited, auditees, auditing, auditor, auditors, audits
Blockchain	32 X 2	Blockchain, Blockchains
Technology	32 X 1	Technology
Fraud	32 X 4	Fraud, frauds, fraudsters, fraudulent
Risk	32 X 2	Risk, risks
Security	32 X 2	Cybersecurity, security
Protection	32 X 1	Protection
Service	32 X 3	Service, services, protection
Quality	32 X 1	Quality

Notes: Information about data files prepared using LitReviewer package. [a] – data set which represents research interest also a study characteristic. [b] - the order of the data matrix in the given data file. [c] – morphemes for a given word pattern.

[5] Available at https://github.com/Kamakshaiah/literature-review/blob/main/textmining.py.

[6] Data set for "audit" word pattern is available at https://github.com/Kamakshaiah/auditing-data-analysis/blob/main/datasets/mining-outputs/audit.xlsx. The master datafile is available at https://github.com/Kamakshaiah/auditing-data-analysis/blob/main/datasets/mining-outputs/audit-master-dataset.csv.

Table 1 shows the details of data files together with respective dimensionality, and word conjugates (morphemes) for a given issue/implication of the study. Word patterns are assumed to be conceptual patterns in later stages of analysis. For instance, the word "Audit" is used for querying a pattern while mining text corpus. Various word conjugates such as "audit", "auditability", "auditable", "audited", "auditees", "auditing", "auditor", "auditors" and "audits" were emerged. These morphemes are assumed to be a component in the text corpus which needs to be tested for univariate dimensionality in later stages of analysis so as to test the homogeneity in the data. The dimensionality refers to the order of data matrix in data file. For instance, the "audit" file has 32 documents arranged in rows and 9 conjugates arranged in columns.

Statistical techniques

At first, descriptive analysis together with a few statistical tests such as D'Agostino and Pearson's normality test (also known as K^2 test), T test, F test, Mood's test and Median tests were used for identifying conceptual patterns and implications. These techniques were performed on term vectors obtained from literature using text mining methodologies (explained in previous sections). A term vector consists of frequencies for a given study pattern against in each of the documents in the corpus. So, the data distribution is a frequency vector of given term/s across sample documents. The measure, which is frequency, is the rate of occurrence of a given word, which is also known as a pattern, in each document. Such data distributions are usually sparse data vectors, and they might be tested using median instead of arithmetic mean as a measure of centrality. The Mood's test, which is basically a median test, is robust for sparse vectors when compared to statistical tests that are not distribution free (Michael and Steven, 1982; Thomas and Janice, 1975).

A few visualization techniques such as box plots and pie charts were used to identify influential documents which are outliers in the collection. Word cloud plot is obtained from master data file to assess content validity, consistency, and reliability. Documents were critically examined for implications based on the observations from box plot and pie charts. In the end a few statistical tests were performed to understand document level concordance and discordance for each of the research challenges. Concordance is the level of agreement or common consent of studies for given implication and the opposite is known as discordance. If the test statistic is statistically significant then it is construed as there is not common consent or level of agreement is different across the studies. At the second stage, multiple regression methods were used to tests the dependencies in between various conceptual patters few of which are also implications. The model for this is as follows:

$$Financial\ auditing\ service = \beta_0 + \beta_1(implication\ 1) + \beta_2(implication\ 2) + \cdots + \beta_n(implication\ n) + \ldots + \beta_1(issue\ 1) + \beta_2(issue\ 2) + \cdots + \beta_n(issue\ n)$$

While this study considers all the word vectors or term vectors as conceptual patterns, these conceptual patterns subjectively ascertained in two categories. Implications are few technologies viz., AI, Blockchain, IoT etc., and coupled with few issues or challenges such as risk, security, fraud, etc. Multiple regression is expected to provide estimates for these implications and challenges which might provide dependencies among concepts.

Data analysis

Content validity, consistency and reliability were assessed using Word Cloud plot. Word cloud also known as tag cloud, is a visual representation of textual data, and is often used to depict keyword metadata. Word cloud generally shows the importance of each tag with font size or color based on the relative importance of the word in given text corpus. Thereafter, the data was mined for each of the study questions to test the hypotheses. Statistical tests, mentioned in the research methodology, were performed on word/term vectors.

Figure 1. Word count plot.
Source: Word-count map created from study data using LitReviewer package.

Figure 1 is the word cloud obtained from sample text corpus. The word cloud used in this study is known as "frequency word cloud". In frequency word cloud the tag size represents the number of times that tag has been applied to research issue. In Figure 1, words such as "big", "data", "audit", "research", are vividly distinct apart from other tags which appear to explain the relationship between "big data analytics" and "financial auditing". This shows that the data is consistent and suitable for study. A careful yet close examination reveals that few tags such as "technology", "information", "algorithm", "security", "fraud", "risk", "quality", "service" appear to be influential in the text corpus. So, the data is not only consistent but also has content validity.

Descriptive analysis

The mining was done for word pattern called "Audit", which is assumed to be one of the essential key words in the text corpus. The data set related to "audit" pattern is available at the article companion repository at GitHub. Table 2 has the summary statistics for the word pattern "Audit".

Table 2
Word pattern related to "Audit"

Word pattern	Mean	Sd[a]	Max	Min	Median	Q1[b]	Q3[c]
Audit	3.03	4.83	18	0	0	0	5.5
auditability	0.03	0.18	1	0	0	0	0
auditable	0.03	0.18	1	0	0	0	0
audited	0.1	0.3	1	0	0	0	0
auditees	0.03	0.18	1	0	0	0	0
auditing	2.84	3.41	18	0	1	1	3.5
auditor	0.48	1.39	7	0	0	0	0
auditors	0.97	2.65	14	0	0	0	1
audits	0.13	0.43	2	0	0	0	0

Source: Obtained from statistical analysis on "Audit" word pattern. Notes: [a] - Standard deviation, [b] – 1st Quartile, [c] - 2nd Quartile, Max – Maximum, Min - Minimum

The word patter for "Audit" is strongly identified with "Audit", "Auditing", "Auditors" and so on. Figure 2 is the box plot for the pattern - "Audit" which highlights conjugates associated with this pattern. The figure shows roughly 16 potential documents as outliers with unique qualities. Outliers in the data related to "Audit" pattern is distinct in Figure 2.

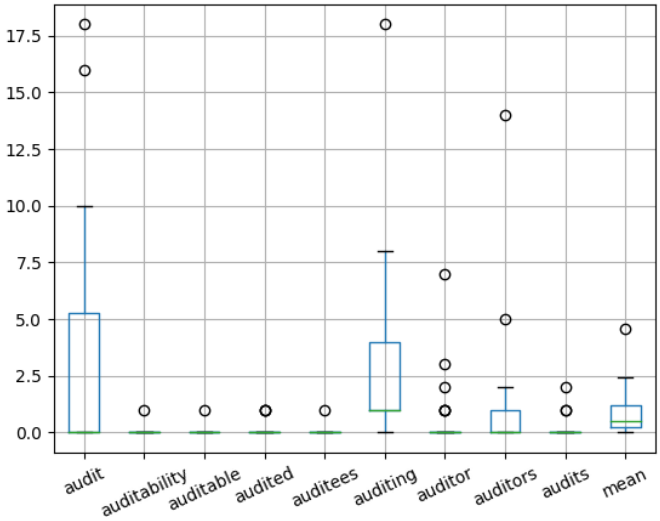

Figure 2. Outliers in "Audit" data.
Source: Obtained from data analysis for "Audit" data set.

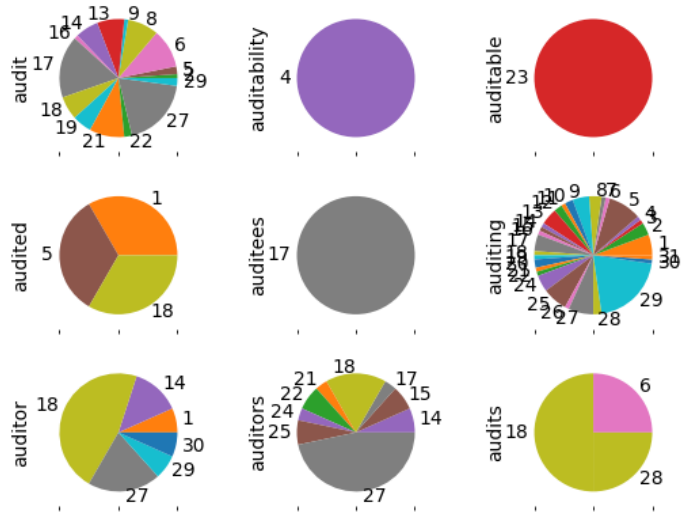

Figure 3. Influential documents in "Audit" data.
Source: Obtained from data analysis for "Audit" data set.

Figure 3 shows the document wise influence for each of the word conjugate associated with word pattern - "Audit". For instance, documents 1, 4, 5, 6, 13, 17, 18, 27, 28, 29, and 30 are distinct compared to the rest of the documents. The information related to these documents was shown in Table 3.

Table 3
Most influential abstracts for "Audit" pattern

Document No.	Authors	Title	Year	Source	Implications
1	Cheong A., Duan H.K., Huang Q., Vasarhelyi M.A., Zhang C.A.	The Rise of Accounting: Making Accounting Information Relevant Again with Exogenous Data	2022	Journal of Emerging Technologies in Accounting	Importance of Accounting
4	[No author name available]	2nd International Conference on Smart Computing and Cyber Security - Strategic Foresight, Security Challenges, and Innovation, SMARTCYBER 2021	2022	Lecture Notes in Networks and Systems	Security challenges and innovation
5	Sharma M.G., Sharma S.M.	Frontiers of Blockchain for Railways	2022	Studies in Big Data	Blockchains for railways
6	Hu K.-H., Chen F.-H., Hsu M.-F., Tzeng G.-H.	Identifying key factors for adopting artificial intelligence-enabled auditing techniques by joint utilization of fuzzy-rough set theory and MRDM technique	2021	Technological and Economic Development of Economy	AI adoption for intelligence enabled auditing
17	Kılıç B.İ.	The effects of big data on forensic accounting practices and education	2020	Contemporary Studies in Economic and Financial Analysis	Forensic accounting practices
18	De Santis F., D'Onza G.	Big data and data analytics in auditing: in search of legitimacy	2020	Meditari Accountancy Research	Legitimacy in auditing
23	Tang J., Karim K.E.	Financial fraud detection and big data analytics – implications on auditors' use of fraud brainstorming session	2019	Managerial Auditing Journal	Financial fraud detection

27	Abrantes P.C., Ferraz F.	Big data applied to tax evasion detection: A systematic review	2017	Proceedings - 2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016	Tax evasion detection
28	Tarek M., Mohamed E.K.A., Hussain M.M., Basuony M.A.K.	The implication of information technology on the audit profession in developing country: Extent of use and perceived importance	2017	International Journal of Accounting and Information Management	Influence of IT on audit profession
29	Alles M., Gray G.L.	Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors	2016	International Journal of Accounting Information Systems	Identification of inhibitors
30	Hegazy M., Tawfik M.	Performance measurement systems in auditing firms: Challenges and other behavioural aspects	2015	Journal of Accounting in Emerging Economies	Performance measurement of auditing firms

This conceptual pattern is found to be associated with few implications such as Technology – Blockchain, AI; Compliance – tax evasion detection, legitimacy in accounting; Challenges – security, fraud detection, forensics, identification of inhibitors for firm; Enablers - performance measurement of firms, influence of IT. These implications show that the concept "audit" is two dimensional in nature. One dimension represents technology enablers, and the other compliance and challenges. This would mean that firms can try big data technologies coupled with Blockchains, AI to detect fraud, try for forensics, identify inhibitors for financial accounting which eventually leads to performance and growth.

Table 4
Statistical tests for "Audit" pattern

Test	Statistic	P Value
k2	305.080	5.66E-67
T	4.789	3.92E-05
F	9.299	2.29E-11
median test	125.903	2.00E-23
Kruskal Wallis	125.74	2.15E-23

Source: From the data analysis.

Table 4 has the statistical tests for the word pattern - "Audit". All the tests are statistically significant which means document wise differences are statistically significant. Studies appear to be different in terms of implications of big data technologies. This means the level of cognizance related to "Audit" does not appear to be equal across the studies. This shows evidence in support of H_1. Document level differences are significant. There is lack of common agreement and the level of opinion is different across the documents. So, big data technologies and their implications are found to be significant for auditing practices.

Table 5
Word pattern related to "Fraud"

Word Conjugate	Mean	Sd	Max	Min	Median	Q1	Q3
Fraud	2.556	1.667	5	1	2	1	2
Frauds	0.111	0.333	1	0	0	0	0
Fraudsters	0.111	0.333	1	0	0	0	0
Fraudulent	0.222	0.667	2	0	0	0	0

Source: document wise word-count for the word 'Fraud'.

"Fraud" is the other distinctly identified pattern in the text corpus. Table 5 shows the summary statistics for the word pattern associated with "Fraud". Conjugates or associated with "fraud" are identified as "fraud", "frauds", "fraudsters" and "fraudulent". Though "fraud" and "frauds" are same, they were retained as it is in the analysis for the purpose of statistics and detailed tabulation. Among all conjugates "fraud" appears to be influential in text corpus. Figure 4 and 5 shows a few potential documents found important for "Fraud".

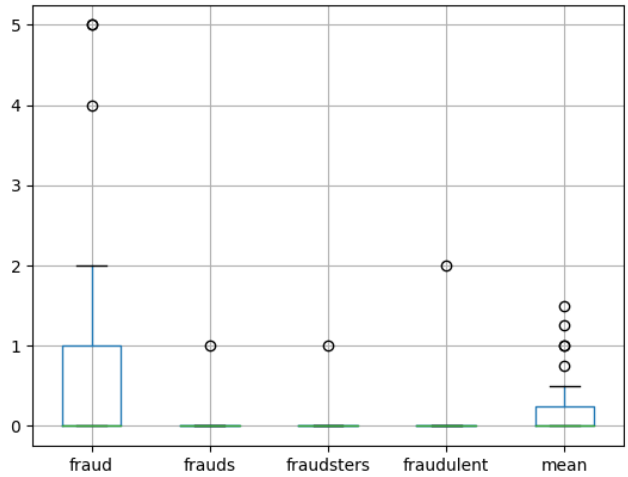

Figure 4. Outliers in "Fraud" data.
Source: Obtained from data analysis for "Fraud" data set.

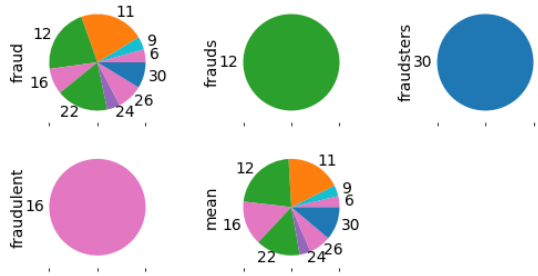

Figure 5. Influential documents in "Fraud" data.
Source: Obtained from data analysis for "Fraud" data set.

Figure 4 and 5 shows roughly 9 outlier documents and 6 influential documents viz., 11, 12, 16, 22, 26, and 30 were highlighted. Conjugates “frauds”, “fraudsters”, and “fraudulent” were identified through a single document. However, “fraud” and “fraudulent” are the most influential conjugates among the identified documents. Table 6 provides the document level information for this pattern.

Table 6
Most influential documents for “Fraud” pattern

Article No.	Authors	Title	Year	Source	Implications
11	[No author name available]	International Conference on Application of Intelligent Systems in Multi-modal Information Analytics, MMIA 2020	2021	Advances in Intelligent Systems and Computing	Intelligent systems; Information analytics
12	Georgakopoulos S.V., Gallos P., Plagianakos V.P.	Using Big Data Analytics to Detect Fraud in Healthcare Provision	2020	Middle East Conference on Biomedical Engineering, MECBME	Fraud in Healthcare
13	Liu R.	Applications of New Technologies to Recognition of Financial Statement Fraud	2020	Proceedings - 2020 International Conference on Computer Communication and Network Security, CCNS 2020	Technologies for recognition financial fraud
16	Handoko B.L., Lindawati A.S.L., Mustapha M.	Application of computer assisted audit techniques in public accounting firm	2020	International Journal of Management	Computer assisted audit techniques for accounting
22	Zhu Y., Huang B.	Summary of research on the application of big data in auditing	2019	Proceedings - 2019 International Conference on Communications, Information System, and Computer Engineering, CISCE 2019	Big data applications
26	Haddara M., Su K.L., Alkayid K., Ali M.	Applications of Big Data Analytics in Financial Auditing- A Study on The Big Four	2018	Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018	Big data analytics for financial auditing
30	Hegazy M., Tawfik M.	Performance measurement systems in auditing firms: Challenges and other behavioural aspects	2015	Journal of Accounting in Emerging Economies	Performance measurement for auditing firms

The pattern "fraud" is strongly associated with implications such as information technologies - AI, big data; challenges – recognition of financial fraud; enabler - performance measurement. These technologies and challenges are found to be significant in healthcare domain. This shows that big data analytics might help while addressing challenges associated with financial fraud. Intelligent (AI) practices found to be influential apart from big data technologies. Performance measurement found to be enabler and a driver for the growth. Recognizing financial fraud appeared as a challenge for "fraud" pattern. Firms can use big data technologies coupled with AI to address financial fraud which eventually such practices might affect firm's performance.

Table 7
Statistical tests for "Fraud" data

Test	Statistic	P Value
k2	153.893	0.00000
T	2.862	0.00748
F	6.483	0.00041
Median test	17.655	0.00052
Kruskal	18.019	0.00044

Source: From the data analysis.

All statistical tests are significant. Both "Mood's test" and the "Median test" shows that document level differences related to patterns associated with "fraud" appear to be statistically significant. This means that there is evidence in support of H_2. Document level differences are significant. There is lack of common agreement and the level of opinion is different across the documents. So, big data technologies and their implications are found to be significant for detecting fraud in financial auditing.

Table 8
Word pattern related to "Security"

Word Conjugate	Mean	Sd	Max	Min	Median	Q1	Q2
Cybersecurity	0.14286	0.37796	1	0	0	0	0
Security	1.71429	1.97605	6	0	1	1	1.5

Source: data analysis on Security data.

"Security" is another issue which is distinct in the text corpus (as ascertained through literature and from word cloud; refer to Figure 1). Table 8 shows the summary statistics and other conjugates of the word pattern associated with "Security". The pattern 'Security' is vividly identified across 6 documents. The conjugates for "Security" are "cybersecurity", and "security". Figure # shows a few outlier documents

and they are 2, 3, 10. While "security" is potentially identified through these documents but only one document is related to "cybersecurity".

Table 9
Most influential documents for "Security" pattern

Article No.	Authors	Title	Year	Source	Implications
2	Lu Y., Qi Y., Qi S., Li Y., Song H., Liu Y.	Say No to Price Discrimination: Decentralized and Automated Incentives for Price Auditing in Ride-Hailing Services	2022	IEEE Transactions on Mobile Computing	Smart price auditing system; decentralized price auditing system; Hyperledger Fabric; Blockchains
3	Jayasuriya D.D., Sims A.	From the abacus to enterprise resource planning: is Blockchain the next big accounting tool?	2022	Accounting, Auditing and Accountability Journal	Blockchain - next big accounting tool
10	Ageeva O., Karp M., Sidorov A.	The Application of Digital Technologies in Financial Reporting and Auditing	2021	Lecture Notes in Networks and Systems	Digital technologies for financial auditing

Interestingly, the pattern "Security" is identified through documents where the interest of the studies appears to be technology related concepts such as "Blockchain" and "digital technologies for auditing". The research interest of document 2 and 3 appears to be "ERP and Blockchain technologies". The study interest of document 10 appears to be "digital technologies" for financial reporting. Overall observation is that digital technologies associated with Blockchains might affect security challenges in auditing practices.

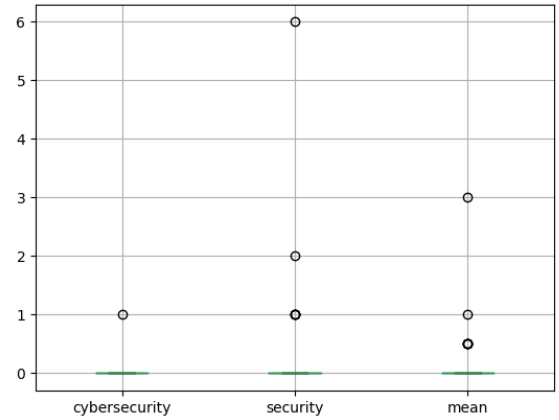

Figure 6. Outliers in "Security" data.
Source: Obtained from data analysis for "Security" data set.

Figure 6 shows the medians for "cybersecurity" and "security". Medians were close to zero. However, this figure is useful for identifying outliers. Though the median difference is distinct, but few documents that were discussed above appear to be distinct outliers. Figure 7 highlights influential documents related to the conjugate "Security". The figure identifies documents 3, 10 as distinct and hence they are influential in the text corpus.

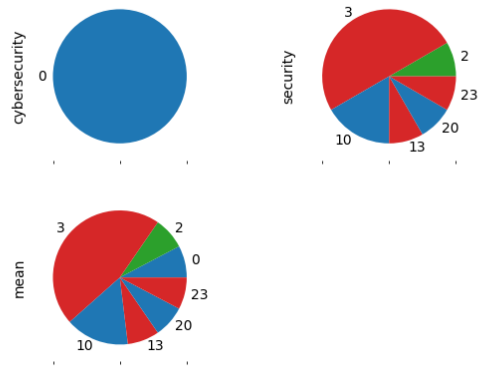

Figure 7. Influential documents in "Security" data.
Source: Obtained from data analysis for "Security" data set.

Table 10
Statistical tests for "Security" data

Test	Statistic	P Value
k2	116.9868	0.0000
T	2.0303	0.0510
F	2.8965	0.0938
Median test	2.5664	0.1092
Kruskal	4.0353	0.0446

Source: From the data analysis.

Table 10 shows the details related to document wise disagreements for "Security". k^2 test is significant. So, the data set for security is non-normal. The Kruskal test is significant showing that the distributions are not identically independent. This means the conjugates 'Cybersecurity' and 'security' are statistically dependent. This shows that these conjugates appear to serve the same purpose as far as big data technologies and auditing practices are concerned. The Mood's test shows that the median difference is insignificant. So, there is same level of cognizance and common consent for 'security' in the literature. Hence, the conceptual pattern "Security" seems to be an insignificant pattern in the study. The evidence for H_3 is in question. However, there is common agreement, and the level of opinion is not different across the documents. So, big data technologies and their implications are not significant though but the findings about requirement for Blockchains are evidenced in the literature.

Table 11
Word pattern related to "Risk"

Word Conjugate	Mean	Sd	Max	Min	Median	Q1	Q2
risk	1.2	1.09545	3	0	1	1	1
risks	1	1.41421	3	0	0	0	2

Source: document wise word-count for the pattern associate with 'risk'.

Risk is another pattern which is vividly distinct in the text corpus. There are two-word patterns in the literature viz., risk and risks. Though these word patterns are not too distinct etymologically but kept distinct for the purpose of analysis of document level opinion differences. Table 11 shows the summary statistics. Figure 8 and Figure 9 shows documents 18 and 27 as influential documents for "risk". The information on these documents is given in Table 12.

Table 12
Most influential documents for word pattern - "Risk"

Article No.	Authors	Title	Year	Source	Implications
27	Abrantes P.C., Ferraz F.	Big data applied to tax evasion detection: A systematic review	2017	Proceedings - 2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016	Big data for tax evasion detection
18	De Santis F., D'Onza G.	Big data and data analytics in auditing: in search of legitimacy	2020	Meditari Accountancy Research	Big data analytics for legitimacy

Documents with titles "Big data applied to tax evasion detection: A systematic review" and "Big data and data analytics in auditing: in search of legitimacy" appear to be influential for this conceptual pattern. The first document emphasizes "tax evasion" and the other "legitimacy". Hence, the conceptual patter "risk" is being observed with "compliance" and tax evasion, legitimacy in financial auditing is found to be implications of big data technologies.

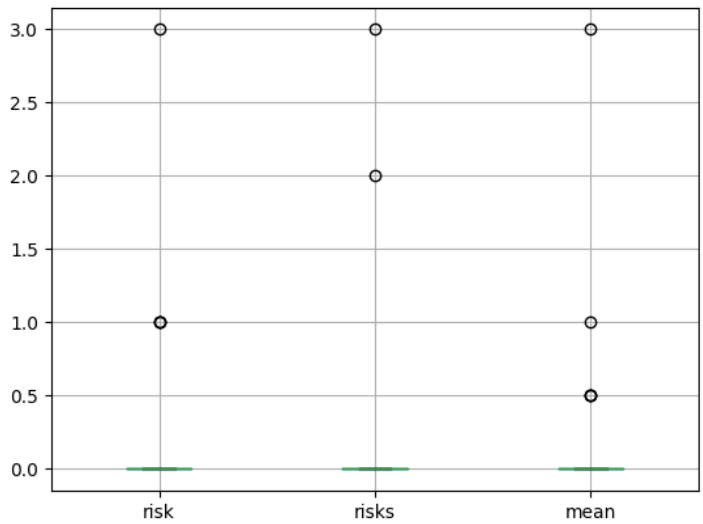

Figure 8. Outliers in "Risk" data.
Source: Obtained from data analysis for "Risk" data set.

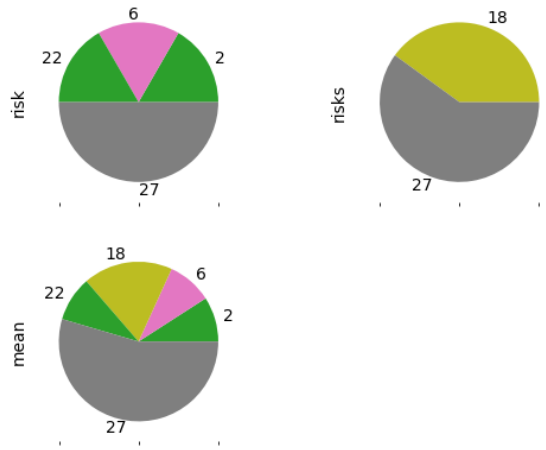

Figure 9. Influential documents in "Risk" data.
Source: Obtained from data analysis for "Risk" data set.

Both box plot and pie charts show documents 27, 18 as influential in the text corpus visually.

Table 13
Statistical tests

Test	Statistic	P Value
k2	79.382	0.000
T	1.729	0.094
F	0.042	0.838
Median test	0.184	0.668
Kruskal	0.614	0.433

Source: From the data analysis.

The data sets are not normally distributed (K2 = 79.382; P value < 0.05). The document level differences for word patterns associated to the issue "Risk" is not statistically significant. The evidence for H_4 is in question. However, there is common agreement, and the level of opinion is not different across the documents. So, big data technologies and their implications are not significant though but the finding about tax evasion and legitimacy in practice are evidenced in the literature.

Multiple regression analysis

Multiple regression is used to test the study model. As it was stated earlier, the primary assumption of the study is to test the impact of big data analytics technologies on financial auditing services. Eventually the influence of implications and challenges on financial auditing services moderated by big data analytics is also tested in conjunction to the primary assumption. The descriptive analysis has shown that Blockchain technology and artificial intelligence were the two fundamental implications of big data analytics. When coming to challenges, big data analytics is found to be effective while addressing risk, security, fraud. In this part of the analysis, these implications together with challenges were analyzed for their influences assuming auditing service as dependent variable and all the implications along with challenges as dependent variables. Table 14 provides the estimates calculated by the multiple regression.

Table 14
Financial auditing vs. study implications

Variable	Estimate [a]	Std. Error [b]	t value [c]	Pr(>|t|) [d]	Effect	Inference
(Intercept)	7.155	2.066	3.464	0.002	Positive	Significant
Bda	-0.352	0.242	-1.452	0.162	Negative	Insignificant
Blockchain	-0.304	0.578	-0.526	0.605	Negative	Insignificant
ai	-12.381	8.126	-1.524	0.143	Negative	Insignificant
fraud	0.083	1.245	0.067	0.947	Positive	Insignificant
risk	-2.431	3.258	-0.746	0.464	Negative	Insignificant
security	-1.185	0.892	-1.328	0.199	Negative	Insignificant
Blockchain:bda	0.014	0.048	0.294	0.772	Positive	Insignificant
bda:ai	6.578	4.247	1.549	0.137	Positive	Insignificant
bda:fraud	-0.170	0.225	-0.755	0.459	Negative	Insignificant
bda:risk	0.529	0.656	0.806	0.430	Positive	Insignificant
bda:security	0.145	0.136	1.070	0.297	Positive	Insignificant

Source: Data analysis on sample study abstracts; Notes: a – beta estimate including intercept and respective slopes for study dependent variables; b – standard error associated with the estimate; c – t statistic for the estimate; d – P Value for inference.

The intercept (7.15519593) is positive and statistically significant. This shows that the influence of implications and challenges are positive and statistically significant (P Value – 0.002). When coming to variable level influences; the effect of Big Data Analytics (BDA) on financial auditing services is negative though insignificant. This means, there is no supporting evidence in the literature that BDA

influences financial auditing services or the statement that the applications of big data analytics for financial auditing is scarce. The other associative technologies i.e., "Blockchains" (-0.304; 0.605), and AI (-12.380; 0.143) also found to be negative but statistically insignificant. This shows that these technologies were referred most insignificantly in those documents where financial auditing was the focus of the study. So, big data analytics and its associative technologies were not found to be important in the financial auditing literature. Interestingly, the moderation effect from Big Data Analytics (BDA) for both Blockchain (0.014; 0.772) and AI (6.578; 0.137) was positive but not significant. So, there is evidence in the literature that big data analytics together with Blockchain and artificial intelligence technologies impacts financial auditing performance.

When coming to challenges; "Fraud" (0.083; 0.947) was found to be an important issue compared to other two i.e., "Risk" (-2.431; 0.464) and "Security" (-1.185; 0.199). Fraud appears to be the focus of the study in the financial auditing literature compared to Risk and Security. Interestingly, the moderation effect from the big data analytics for Fraud was found to be negative and positive for the rest of the challenges i.e., Risk and Security. So, Big Data Analytics (BDA) appears to be important in those studies where the focus is brought on both Risk and Security but on Fraud. Figure 10 adds visualization to the above interpretation and to the statistics in Table 14.

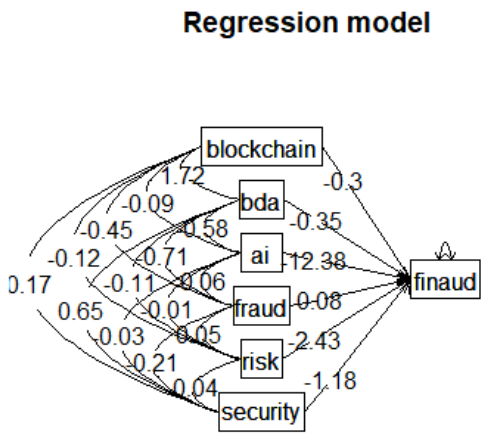

Figure 10. LM plot for study model.

Conclusion

Big data technologies are closely connected to the participants of financial auditing services. The analysis shows the evidence in favor of big data technologies and their impact on financial auditing practices. For instance, document level differences related to patterns associated with "Audit", "Fraud", "Risk" appear to be statistically significant and the level of cognizance in the literature is not equal. This shows evidence in support of big data technologies and its importance to "financial auditing services" in general and for fraud detection and risk mitigation in particular. The ideas related to these concepts are divergent in the literature. Interestingly, the pattern "Security" is identified with those studies where the research interest appears to be "information technology". Though the conjugates for this pattern appeared statistically insignificant but this pattern is potentially identified with few implications such as digital technology in general and Blockchain based technologies such as Hyperledger in particular. The support for the idea of security is not similar in literature.

Implications & challenges

- There is sufficient emphasis on digital technologies and information systems in financial auditing literature.
- Technologies such as Blockchain, Artificial Intelligence (AI) were found to be main implications of Big data analytics (BDPT).
- These technologies are found to be useful for tax evasion detection and legitimacy in accounting.
- A few challenges such as Fraud, Risk, Security were identified in the literature.
- Big data technologies were also found to be associated with a few enablers such as performance measurement and growth.
- The overall impacts in the study model are positive and statistically significant. This means Big Data Analytics together with its implications and challenges appear to influence financial auditing practices.
- Applications of Big Data Analytics (BDA), Blockchain technologies and Artificial Intelligence (AI) were not found to be the focus of literature individually, but moderation effects in the analysis shows that the BDA found to be important when the main goal of the study was either Blockchain technology or AI.

- Fraud was found to be highly emphatic in the literature but there are very few studies with Fraud being associated with BDPT. The challenges such as Risk and Security were not found to be the specific interest in the literature, but these challenges were found to be largely associated with BDPT.

References

Abdelwahed, A. S., Abu-Musa, A. A., Moubarak, H., & Badawy, H. A. (2024). The use of big data and analytics in external auditing: Does audit firm size matter? Evidence from a developing country. South African Journal of Accounting Research, 1-33. http://dx.doi.org/10.1080/10291954.2023.2279751

Abu Huson, Y., Sierra-García, L., & Garcia-Benau, M. A. (2024). A bibliometric review of information technology, artificial intelligence, and blockchain on auditing. Total Quality Management & Business Excellence, 35(1-2), 91-113. https://doi.org/10.1080/14783363.2023.2256260

Ahmadi, S. (2024). A Comprehensive Study on Integration of Big Data and AI in Financial Industry and its Effect on Present and Future Opportunities. International Journal of Current Science Research and Review, 7(01), 66-74. https://doi.org/10.47191/ijcsrr/v7-i1-07

Ahmad, A. Y. A. B., Abusaimeh, H., Rababah, A., Alqsass, M., Al-Olima, N., & Hamdan, M. (2024). Assessment of effects in advances of accounting technologies on quality financial reports in Jordanian public sector. Uncertain Supply Chain Management, 12(1), 133-142. https://doi.org/10.5267/j.uscm.2023.10.011

Alharasis, E. E., Alkhwaldi, A. F., & Hussainey, K. (2024). Key audit matters and auditing quality in the era of COVID-19 pandemic: the case of Jordan. International Journal of Law and Management. https://doi.org/10.1108/ijlma-11-2023-0248

Alles, M., & Gray, G. L. (2016). Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors. International Journal of Accounting Information Systems, 22, 44-59. https://doi.org/10.1016/j.accinf.2016.07.004

Almasria, N., Airout, R. M., Samara, A. I., Saadat, M., & Jrairah, T. S. (2021). The role of accounting information systems in enhancing the quality of external audit procedures. Journal of management Information and Decision Sciences, 24(7), 1-23.

Alrashidi, M., Almutairi, A., & Zraqat, O. (2022). The impact of big data analytics on audit procedures: Evidence from the Middle East. The Journal of Asian Finance, Economics and Business, 9(2), 93-102.

Alzahrane, M. A. (2024). Insights from forensic accounting educators and practitioners within the KSA context regarding the optimal forensic accounting skills set: an implication on the

socioeconomic development. Journal of Business and Socio-economic Development, 4(1), 66-80. https://doi.org/10.1108/jbsed-08-2022-0082

Amir H. G., Chen, F. and Abualigah, L. (2022). "Machine Learning Technologies for Big Data Analytics" Electronics 11, no. 3: 421. https://doi.org/10.3390/electronics11030421

Anitha, C., Tandon, S., Vamsikrishna, M. M. A., Chauhan, M., & Bajaj, R. (2024). Role of IoT Intelligence System & Big Data Management to Control Flood Data. International Journal of Intelligent Systems and Applications in Engineering, 12(15s), 455–463. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4769

Anwar, S., Shukla, V. K., Rao, S. S., Sharma, B. K., & Sharma, P. (2019). Framework for financial auditing process through blockchain technology, using identity based cryptography. In 2019 Sixth HCT Information Technology Trends (ITT) (pp. 099-103). IEEE. https://doi.org/10.1109/itt48889.2019.9075120

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory, 36(4), 1-27. https://doi.org/10.2308/ajpt-51684

Azzam, M. E. A. Y., Alsayed, M. S. H., Alsultan, A., & Hassanein, A. (2024). How big data features drive financial accounting and firm sustainability in the energy industry. Journal of Financial Reporting and Accounting, 22(1), 29-51. https://doi.org/10.1108/jfra-03-2023-0125

Bejjar, M. A., & Siala, Y. (2024). The Impact of Blockchain Technology on the Financial Audit. In Impact of Digitalization on Reporting, Tax Avoidance, Accounting, and Green Finance (pp. 272-300). IGI Global. https://doi.org/10.4018/979-8-3693-1678-8.ch012

Bramer, W. M., De Jonge, G. B., Rethlefsen, M. L., Mast, F., & Kleijnen, J. (2018). A systematic approach to searching: an efficient and complete method to develop literature searches. Journal of the Medical Library Association: JMLA, 106(4), 531. https://doi.org/10.5195/jmla.2018.283

Bonyuet, D. (2020). Overview and impact of blockchain on auditing. International Journal of Digital Accounting Research, 20, 31-43. https://doi.org/10.4192/1577-8517-v20_2

Bose, S., Dey, S. K., & Bhattacharjee, S. (2023). Big data, data analytics and artificial intelligence in accounting: An overview. Handbook of big data research methods, 32-51. https://doi.org/10.4337/9781800888555.00007

Boynton, J., Glanville, J., McDaid, D., & Lefebvre, C. (1998). Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. Journal of information science, 24(3), 137-154. https://doi.org/10.1177/016555159802400301

Brender, N., Gauthier, M., Morin, J. H., & Salihi, A. (2024). Three lines model paradigm shift: a blockchain-based control framework. Journal of applied accounting research, 25(1), 149-170. https://doi.org/10.1108/jaar-06-2022-0143

Chu, M. K., & Yong, K. O. (2021). Big data analytics for business intelligence in accounting and audit. Open Journal of Social Sciences, 9(9), 42-52. https://doi.org/10.4236/jss.2021.99004

De Santis, F., & D'Onza, G. (2021). Big data and data analytics in auditing: in search of legitimacy. Meditari Accountancy Research, 29(5), 1088-1112. https://doi.org/10.1108/medar-03-2020-0838

Dagilienė, L., & Klovienė, L. (2019). Motivation to use big data and big data analytics in external auditing. Managerial Auditing Journal. https://doi.org/10.1108/maj-01-2018-1773

Desplebin, O., Lux, G., & Petit, N. (2021). To be or not to be: Blockchain and the future of accounting and auditing. Accounting Perspectives, 20(4), 743-769. https://doi.org/10.1111/1911-3838.12265

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. Academy of Management Journal, 59(5), 1493-1507. https://doi.org/10.5465/amj.2016.4005

Grabski, S. V., Leech, S. A., & Schmidt, P. J. (2011). A review of ERP research: A future agenda for accounting information systems. Journal of information systems, 25(1), 37-78. https://doi.org/10.2308/jis.2011.25.1.37

Hasan, A. R. (2021). Artificial Intelligence (AI) in accounting & auditing: A Literature review. Open Journal of Business and Management, 10(1), 440-465. https://doi.org/10.4236/ojbm.2022.101026

Hashem, I. A. T., et al (2016). The role of big data in smart city. International Journal of information management, 36(5), 748-758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Hezam, Y. A., Anthonysamy, L., & Suppiah, S. D. K. (2023). Big data analytics and auditing: A review and synthesis of literature. Emerging Science Journal, 7(2), 629-642. https://doi.org/10.28991/esj-2023-07-02-023

Hima, Z., Yusoff, W. S., & binti Saraih, U. N. (2024). Big Data and Audit Quality: Theoretical Study. South Asian Journal of Social Sciences and Humanities, 5(1), 144-160. https://doi.org/10.48165/sajssh.2024.5109

Hosseinpour, M., & Jans, M. (2024). Auditors' Categorization of Process Deviations. Journal of Information Systems, 38(1), 67-89. https://doi.org/10.2308/isys-2023-051

Ioannou, S., Wojcik, D., & Pažitka, V. (2021). Financial centre bias in sub-sovereign credit ratings. Journal of International Financial Markets, Institutions and Money, 70, 101261. https://doi.org/10.1016/j.intfin.2020.101261

Ionescu, B., Ionescu, D., Gadea, C., Solomon, B., & Trifan, M. (2014, July). An architecture and methods for big data analysis. In International Workshop Soft Computing Applications (pp. 491-514). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-18296-4_39

Jan, C. L. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. Sustainability, 10(2), 513. https://doi.org/10.3390/su10020513

Jin, X., Wah, B. W., Cheng, X., Wang, Y. (2015). Significance and Challenges of Big Data Research, Big Data Research, Volume 2, Issue 2, Pages 59-64. https://doi.org/10.1016/j.bdr.2015.01.006

Johnson, J.S., Friend, S.B. and Lee, H.S. (2017), Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process. Journal of Product Innovation Management, 34: 640-658. https://doi.org/10.1111/jpim.12397.

Kanaparthi, V. (2024). Exploring the Impact of Blockchain, AI, and ML on Financial Accounting Efficiency and Transformation. arXiv preprint arXiv:2401.15715. https://doi.org/10.1007/978-981-97-1488-9_27

Kazemi, M., Majumdar, R., Salamati, M., Soudjani, S., & Wooding, B. (2024). Data-driven abstraction-based control synthesis. Nonlinear Analysis: Hybrid Systems, 52, 101467. https://doi.org/10.1016/j.nahs.2024.101467

Khan, N., et al (2014). "Big Data: Survey, Technologies, Opportunities, and Challenges", The Scientific World Journal, vol. 2014, Article ID 712826, 18 pages, 2014. https://doi.org/10.1155/2014/712826

Knechel, W. R., & Salterio, S. (2016). Auditing: Assurance and risk. Routledge. https://doi.org/10.4324/9781315531731

Liu, W., Wang, J., Jia, F., & Choi, T. M. (2022). Blockchain announcements and stock value: a technology management perspective. International Journal of Operations & Production Management, 42(5), 713-742. https://doi.org/10.1108/ijopm-08-2021-0534

Malsch, B., & O'Dwyer, B. (2021). New directions in auditing research: Conceptual repair, technological disruption (s), local professional governance and the battle for inclusivity. European Accounting Review, 30(3), 439-444. https://doi.org/10.1080/09638180.2021.1924813

Manita, R., Elommal, N., Baudier, P., & Hikkerova, L. (2020). The digital transformation of external audit and its impact on corporate governance. Technological Forecasting and Social Change, 150, 119751. https://doi.org/10.1016/j.techfore.2019.119751

Musunuru, K. (2021). litreviewer: A Python Package for Review of Literature (RoL). Available at SSRN 3856547. https://doi.org/10.2139/ssrn.3856547

McNeely, C.L. and Hahm, J.-o. (2014), Big (Data) Bang. Review of Policy Research, 31: 304-310. https://doi.org/10.1111/ropr.12082

Nguyen, N., Dang-Van, T., Vo-Thanh, T., Do, H. N., & Pervan, S. (2024). Digitalization strategy adoption: The roles of key stakeholders, big data organizational culture, and leader commitment. International Journal of Hospitality Management, 117, 103643. https://doi.org/10.1016/j.ijhm.2023.103643

Ochuba, N. A., Amoo, O. O., Okafor, E. S., Akinrinola, O., & Usman, F. O. (2024). Strategies for leveraging big data and analytics for business development: a comprehensive review across sectors. Computer Science & IT Research Journal, 5(3), 562-575. https://doi.org/10.51594/csitrj.v5i3.861

Olabanji, S. O., Oladoyinbo, O. B., Asonze, C. U., Oladoyinbo, T. O., Ajayi, S. A., & Olaniyi, O. O. (2024). Effect of Adopting AI to Explore Big Data on Personally Identifiable Information (PII) for Financial and Economic Data Transformation. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4739227

Oluwagbade, O. I., Boluwaji, O. D., Azeez, O. A., & Njengo, L. M. (2024). Challenges and Opportunities of Implementing Artificial Intelligence in Auditing Practices: A Case Study of Nigerian Accounting Firms. Asian Journal of Economics, Business and Accounting, 24(1), 32–45. https://doi.org/10.9734/ajeba/2024/v24i11210

Patel, S., & Shah, M. (2023). A Comprehensive Study on Implementing Big Data in the Auditing Industry. Annals of Data Science, 10(3), 657-677. https://doi.org/10.1007/s40745-022-00430-8

Peters, G. T., & Onamariwari, O. B. (2024). Integrating it governance into big data analytics and its relationship with audit reporting lag: a conceptual framework. Advance Journal of Management, Accounting and Finance, 9(1), 24-34.

Pothumsetty, R. (2020). Implementation of Artificial Intelligence and Machine learning in Financial services. International Research Journal of Engineering and Technology, 7(03). https://doi.org/10.33564/ijeast.2020.v04i12.019

Rahman, S. F., & Irwansyah, I. (2024). The role of big data in audit quality and fraud disclosure. In Proceeding International Conference on Accounting and Finance (pp. 467-476).

Rana, T., Svanberg, J., Öhman, P., Lowe, A. (2023). Introduction: Analytics in Accounting and Auditing. In: Rana, T., Svanberg, J., Öhman, P., Lowe, A. (eds) Handbook of Big Data and Analytics in Accounting and Auditing. Springer, Singapore. https://doi.org/10.1007/978-981-19-4460-4_1.

Rijanto, A. (2024). Blockchain technology roles to overcome accounting, accountability and assurance barriers in supply chain finance. Asian Review of Accounting. https://doi.org/10.1108/ara-03-2023-0090

Sangster, A., Leech, S. A., & Grabski, S. (2009). ERP implementations and their impact upon management accountants. JISTEM-Journal of Information Systems and Technology Management, 6, 125-142. https://doi.org/10.4301/s1807-17752009000200001

Sanoran, K., & Ruangprapun, J. (2023). Initial implementation of data analytics and audit process management. Sustainability, 15(3), 1766. https://doi.org/10.3390/su15031766

Schildt, H. (2017) Big data and organizational design – the brave new world of algorithmic management and computer augmented transparency, Innovation, 19:1, 23-30, DOI: 10.1080/14479338.2016.1252043

Sihombing, R. P., Narsa, I. M., & Harymawan, I. (2023). Big data analytics and auditor judgment: an experimental study. Accounting Research Journal, 36(2/3), 201-216. https://doi.org/10.1108/arj-08-2022-0187

Smith, S. S., (2018). Blockchain augmented audit – benefits and challenges for accounting professionals. The Journal of Theoretical Accounting Research; New Rochelle Vol. 14, Iss. 1, 117-137.

Sun, Y., Li, J., Lu, M., & Guo, Z. (2024). Study of the Impact of the Big Data Era on Accounting and Auditing. arXiv preprint arXiv:2403.07180. https://doi.org/10.54097/0fa7xk79

Ülkü, M. A., & Mansouri, B. (2024). Supply Chain Analytics: Overview, Emerging Issues, and Research Outlook. The Palgrave Handbook of Supply Chain Management, 1275-1299. https://doi.org/10.1007/978-3-031-19884-7_80

UoB, (n.d.). How to do a Literature Search: Using keywords. Retrieved from https://library.bath.ac.uk/literaturesearch/using-keywords.

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

Vitali, S., & Giuliani, M. (2024). Emerging digital technologies and auditing firms: Opportunities and challenges. International Journal of Accounting Information Systems, 53, 100676. https://doi.org/10.1016/j.accinf.2024.100676

Xie, M. (2019, April). Development of artificial intelligence and effects on financial system. In Journal of Physics: Conference Series (Vol. 1187, No. 3, p. 032084). IOP Publishing. https://doi.org/10.1088/1742-6596/1187/3/032084

Wassie, F. A., & Lakatos, L. P. (2024). Artificial intelligence and the future of the internal audit function. Humanities and Social Sciences Communications, 11(1), 1-13. https://doi.org/10.1057/s41599-024-02905-w