Sciences Methods and Technologies International Journal (SciMeTech)

ISSN: 3085-5284, journal homepage: www.scimetech.com

Artificial Intelligence and Audit Effectiveness in Global Accounting: A Big Four Analysis

Saad Saadouni 1, Souad Habbani 2

1. *LIREFIMO Laboratory, Faculty of Law, Economics and Social Sciences, University Sidi Mohammed Ben Abdellah, Fez-Morocco.* Saad.saadouni@usmba.ac.ma
2. *LIREFIMO Laboratory, Faculty of Law, Economics and Social Sciences, University Sidi Mohammed Ben Abdellah, Fez-Morocco.* Souad.habbani@usmba.ac.ma

How to cite:

Saad Saadouni , Souad Habbani, "Artificial Intelligence and Audit Effectiveness in Global Accounting: A Big Four Analysis", Sciences Methods and Technologies International Journal (SciMeTech), Vol 2, Issue 2, p 133-154

Keywords:
Big Four,
Audit effectiveness,
Financial risk,
AI adoption,
Compliance violations,
Fraud detection,
KPMG,
Deloitte,
PwC,
Ernst & Young

Abstract

The Big Four accounting firms (Deloitte, PwC, Ernst & Young, and KPMG) face unprecedented challenges in financial risk management, compliance, and fraud detection amid rapid technological transformation and regulatory evolution. This study examines the relationship between AI adoption, employee workload, and audit effectiveness across Big Four firms from 2020-2025, analyzing high-risk cases, compliance violations, and fraud detection patterns across four major industries. We analyzed 100 audit engagement records spanning 2020-2025, employing descriptive statistics, correlation analysis, trend analysis, and comparative performance evaluation. The dataset encompasses 12 key variables including audit engagements, risk cases, compliance violations, fraud detection, AI adoption, employee workload, and satisfaction metrics. Our findings reveal significant variation in high-risk case management across firms, with mean high-risk cases of 277.73 (SD=135.74). Healthcare sector accounted for the highest audit engagements (68,004), followed by Technology (80,069). AI adoption showed mixed correlation with audit effectiveness. Revenue impact averaged 272.54 $M per engagement, with employee workload averaging 60.25 hours.

The study demonstrates that audit effectiveness is multifactorial, influenced by AI adoption, workload management, and industry-specific risk profiles. Strategic AI integration, combined with optimal workload distribution, significantly enhances fraud detection and compliance outcomes.

I. Introduction

The global auditing profession has entered an unprecedented era of transformation, driven by technological innovation, regulatory complexity, and evolving stakeholder expectations. The integration of artificial intelligence (AI) and advanced analytics into audit practices represents one of the most significant paradigm shifts since the establishment of modern auditing standards in the mid-twentieth century. This technological revolution coincides with heightened regulatory scrutiny following high-profile audit failures, increased compliance complexity across multiple jurisdictions, and growing demands for real-time assurance and forward-looking risk assessments. The Big Four accounting firms—Deloitte, PwC, Ernst & Young, and KPMG—collectively audit over 80% of publicly traded companies worldwide and serve as gatekeepers of financial market integrity, making their adoption and implementation of AI technologies critically important for global financial stability and investor confidence. Understanding how AI adoption influences audit effectiveness, fraud detection capabilities, and overall risk management performance has become essential for policymakers, regulators, audit committees, and the broader business community seeking to navigate an increasingly complex and interconnected global economy.

The relationship between technology adoption and audit quality has been a subject of scholarly inquiry for decades, with early research focusing on computerized audit tools and data analytics capabilities. The audit profession's technological evolution began with basic spreadsheet applications in the 1980s, progressed through enterprise resource planning system integration in the 1990s, expanded with continuous auditing concepts in the 2000s, and now encompasses artificial intelligence, machine learning, blockchain verification, and robotic process automation in the 2020s. However, the pace and scale of AI adoption during the 2020-2025 period represent a qualitative departure from previous technological transitions. Unlike earlier innovations that primarily enhanced efficiency without fundamentally altering audit methodology, AI technologies promise to transform core audit processes including risk assessment, analytical procedures, substantive testing, and fraud detection through capabilities such as natural language processing for contract analysis, anomaly detection algorithms for transaction testing, predictive analytics for risk identification, and computer vision for physical inventory verification. The COVID-19 pandemic accelerated this transformation by necessitating remote audit procedures, highlighting the limitations of traditional sampling-based approaches, and demonstrating technology's potential to maintain audit quality under unprecedented constraints. Big Four firms invested substantially in proprietary AI platforms during this period, with Deloitte developing Argus for anomaly detection, PwC launching GL.ai for general ledger analysis, Ernst & Young implementing Canvas for workflow integration, and KPMG deploying Clara across the audit lifecycle, collectively representing billions of dollars in research and development expenditure.

Despite significant technological investment and widespread enthusiasm regarding AI's transformative potential, empirical evidence concerning actual performance improvements remains limited and often contradictory. Several studies document positive associations between technology adoption and audit quality proxies, with researchers finding that data analytics usage correlates with enhanced fraud detection rates, reduced audit deficiencies, and improved financial reporting quality. For instance, Cao et al. (2015) [11], demonstrated that Big Data analytics enable auditors to examine entire transaction populations rather than samples, potentially identifying anomalies that traditional methods would miss. Similarly, Kokina and Davenport (2017) [35] argued that AI and machine learning applications could revolutionize audit methodology by automating routine procedures and enabling auditors to focus on complex judgment areas requiring professional skepticism and business understanding. Brown-Liburd et al. (2015) found that blockchain technology integration could provide continuous assurance capabilities, fundamentally altering the temporal nature of auditing from periodic examinations to real-time verification. These optimistic perspectives suggest that AI adoption should substantially enhance audit effectiveness across multiple dimensions including risk assessment accuracy, evidence evaluation comprehensiveness, fraud detection capability, and overall quality consistency.

However, other research streams identify significant challenges, limitations, and potential risks associated with AI adoption in professional services contexts. Eulerich et al. (2021) [26] cautioned that algorithm opacity creates interpretability challenges, potentially undermining audit trail transparency and professional accountability. They argued that auditors must understand how AI systems reach conclusions to exercise appropriate professional skepticism and defend findings to regulators and stakeholders. Munoko et al. (2020) [41] highlighted ethical implications of AI usage in auditing, including algorithmic bias concerns, privacy considerations, and the fundamental question of whether machines can exercise professional judgment. Commerford et al. (2022) [17] found that auditors sometimes exhibit overreliance on AI outputs, accepting algorithm recommendations without sufficient critical evaluation, thereby potentially introducing new audit risks even while addressing traditional weaknesses. The tension between AI's promise and practical implementation challenges reflects broader debates within information systems research regarding technology acceptance, user resistance, organizational change management, and the complex process through which technological capabilities translate into performance improvements.

Technology proponents argue that AI adoption should alleviate workload pressures by automating routine procedures, accelerating data analysis, and enabling more efficient audit execution. If AI systems can independently perform tasks such as journal entry testing, account reconciliation verification, and basic analytical procedures, auditors should theoretically have more time available for complex judgment areas requiring human expertise. This efficiency hypothesis suggests that AI adoption and workload management represent complementary strategies for audit quality enhancement, with technology enabling firms to maintain quality standards even under challenging time constraints. However, critics note that technology adoption itself creates additional workload through learning curves, system maintenance requirements, troubleshooting needs, and the necessity of validating AI outputs before relying on them for audit evidence. Furthermore, if AI enables more comprehensive testing, firms might expand audit scope rather than reducing hours, leaving workload levels unchanged while potentially improving quality through enhanced coverage. The empirical relationship between AI adoption, workload levels, and audit effectiveness therefore remains an open question requiring rigorous investigation.

The Big Four firms themselves present interesting variation in strategic approaches, organizational cultures, technology platforms, and performance outcomes despite operating under similar regulatory frameworks and serving comparable client portfolios. Each firm has developed proprietary AI platforms reflecting different technological philosophies and implementation strategies. Deloitte's Argus platform emphasizes comprehensive data analytics and anomaly detection across entire transaction populations, reflecting a philosophy of exhaustive testing enabled by computational power. PwC's

GL.ai focuses on natural language processing for contract and document analysis, recognizing that much audit evidence exists in unstructured text requiring interpretation. Ernst & Young's Canvas platform prioritizes workflow integration and global consistency, seeking to standardize audit methodology across offices and engagement teams worldwide. KPMG's Clara emphasizes auditor-friendly interfaces and human-AI collaboration, recognizing that technology acceptance depends critically on user experience and seamless integration with existing work practices. These strategic differences suggest that identical technological capabilities may yield different performance outcomes depending on implementation approaches, organizational support systems, training investments, and cultural factors influencing technology acceptance and utilization patterns.

Regulatory developments during the 2020-2025 period further shaped the context for audit quality evolution. The Public Company Accounting Oversight Board increased inspection intensity and expanded focus areas to include technology usage, audit firm quality control systems, and root cause analysis of identified deficiencies. Inspection reports documented deficiency rates ranging from 18% to 42% across Big Four firms, generating reputational concerns and regulatory pressure for quality improvements. The Securities and Exchange Commission pursued enforcement actions against all Big Four firms for various audit failures, including cases involving revenue recognition errors, inventory valuation deficiencies, and insufficient fraud risk assessment. International regulatory bodies including the Financial Reporting Council in the United Kingdom and the Financial Services Agency in Japan similarly intensified oversight and imposed sanctions for quality shortfalls. This heightened regulatory environment created strong external pressures for audit quality enhancement while simultaneously constraining the types of AI applications permissible without explicit regulatory approval, as auditors remained uncertain about regulators' willingness to accept algorithm-driven evidence without traditional confirmation procedures.

The period from 2020 to 2025 also witnessed significant economic disruptions that influenced audit risk profiles and quality dynamics. The COVID-19 pandemic created unprecedented uncertainty regarding asset valuations, going concern assessments, revenue recognition for disrupted contracts, and inventory obsolescence in sectors experiencing demand collapses. Government stimulus programs generated unusual transactions requiring audit attention, including Paycheck Protection Program loan accounting, employee retention credit claims, and various subsidy programs with complex eligibility criteria. Supply chain disruptions complicated inventory audits and raised questions about controls over physical assets when traditional observation procedures proved impossible under lockdown restrictions. The subsequent inflation surge created fair value measurement challenges, impairment testing complications, and questions about historical cost principle applicability in rapidly changing price environments. Geopolitical tensions including Ukraine conflict, US-China trade disputes, and Middle East instability generated sanctions compliance issues, asset impairment considerations, and going concern questions for multinational enterprises. These economic disruptions created a high-risk audit environment testing audit firms' ability to maintain quality under challenging circumstances while simultaneously accelerating technology adoption as traditional procedures proved inadequate.

Table 1: Literature Summary for Audit Quality and Technology Adoption

Author(s)	Country/Region	Period	Method	Key Findings
Cao et al. (2015) [11]	USA	2010-2013	Survey & Case Study	Big Data analytics enable population testing vs. sampling; enhances anomaly detection
Kokina & Davenport (2017) [35]	Global	Conceptual	Literature Review	AI and ML can automate routine tasks; allows focus on judgment areas requiring skepticism
Brown-Liburd et al. (2015)	USA	Conceptual	Theoretical Framework	Blockchain enables continuous assurance; fundamentally alters temporal audit nature
Eulerich et al. (2021) [26]	Germany	2018-2020	Experimental	Algorithm opacity creates interpretability challenges; undermines audit trail transparency
Munoko et al. (2020) [41]	Multi-country	Conceptual	Ethical Analysis	AI raises bias concerns, privacy issues; questions whether machines can exercise judgment

Commerford et al. (2022) [17]	USA	2019-2021	Experimental	Auditors exhibit overreliance on AI outputs; accept recommendations without critical evaluation
López & Peters (2012) [38]	USA	2000-2009	Archival Analysis	Workload compression significantly impairs quality through reduced evidence evaluation depth
DeFond & Zhang (2014) [22]	Global	1980-2012	Literature Review	Audit quality is multidimensional: competence, independence, skepticism, risk assessment
DeAngelo (1981) [21]	USA	Theoretical	Economic Model	Larger firms have greater reputational capital, incentivizing quality maintenance
Francis & Wang (2008) [30]	42 countries	1994-2004	Archival	Big 4 audits show higher quality through lower discretionary accruals globally
Lawrence et al. (2011) [37]	USA	2004-2007	Archival	Office-level characteristics significantly influence quality; variation exists within Big 4
Choi et al. (2010) [15]	South Korea	2001-2005	Archival	Industry specialization enhances audit quality through expertise accumulation
Alles & Gray (2021) [3]	USA	2020	Conceptual	COVID-19 highlighted traditional audit limitations; demonstrated technology potential
Austin et al. (2021) [4]	USA	2017-2020	Survey & Interview	Balanced workload distribution enhances efficiency and effectiveness simultaneously
Guénin-Paracini et al. (2014) [31]	France	2010-2012	Qualitative	Technology can alleviate routine burdens; enables focus on complex judgment areas
ACFE (2022) [1]	Global	2020-2022	Survey	Organizations lose 5% revenue to fraud; external audits detect only 4% of frauds
Perols & Lougee (2011) [43]	USA	2000-2008	Archival	Predictive models improve fraud risk assessment accuracy over traditional methods
Carcello & Nagy (2004) [12]	USA	1990-1999	Archival	Industry specialization reduces fraud risk through superior risk assessment
Ferguson et al. (2003) [27]	Australia	1996-2000	Archival	Industry expertise enables efficient resource allocation and superior risk assessment
Beattie et al. (1999) [6]	UK	1997-1998	Survey	Client satisfaction includes communication, business understanding beyond technical quality
Carcello et al. (1992) [13]	USA	1988-1990	Survey	Communication quality and expertise drive satisfaction more than client accommodation

Moore et al. (2006) [40]	USA	Conceptual	Theoretical	Excessive satisfaction focus risks compromising skepticism and independence
Zhang et al. (2022) [50]	China	2015-2020	Archival	AI and blockchain technologies improve accounting accuracy and fraud detection
Moffitt et al. (2018) [39]	USA	Conceptual	Framework	Optimal outcomes emerge from human-AI collaboration rather than full automation
Albitar et al. (2020) [2]	Global	2020	Conceptual	COVID-19 intensified workload through remote work, client distress, regulatory changes
Watkins et al. (2004) [49]	USA	Theoretical	Literature Synthesis	Healthcare presents complex regulatory environment requiring specialized expertise
Brazel et al. (2009) [9].	USA	2005-2007	Experimental	Technology sector requires expertise in intangible valuation and revenue recognition
Cohen et al. (2011) [16]	USA	2000-2008	Survey & Case	Finance sector faces extensive regulation and fair value measurement challenges
Hammersley (2006) [33]	USA	1999-2002	Archival	Retail presents inventory valuation issues and supply chain complexity
PCAOB (2023) [42]	USA	2022	Inspection Report	Deficiency rates range 18-42% across Big 4; increased technology usage scrutiny

Source: Authors

Against this rich contextual background, systematic empirical investigation of relationships among AI adoption, employee workload, audit effectiveness, client satisfaction, and industry-specific risk profiles becomes essential for understanding contemporary audit quality determinants. While substantial literature exists examining individual factors in isolation, comprehensive studies integrating these multiple dimensions within a unified analytical framework remain scarce. Most existing research relies on survey methodologies capturing auditor perceptions rather than objective performance metrics, limiting causal inference and generalizability. Studies comparing performance across all Big Four firms simultaneously are particularly rare, with most research focusing on single firms, specific industries, or broad audit market segments without distinguishing among leading competitors. The 2020-2025 period represents a particularly important yet understudied timeframe, encompassing pandemic impacts, accelerated technology adoption, regulatory intensification, and economic disruption that collectively create a unique context for audit quality research.

This study addresses these gaps by analyzing comprehensive audit engagement data across all Big Four accounting firms during the 2020-2025 period, examining relationships among AI adoption, employee workload, audit effectiveness scores, client satisfaction measures, compliance violation rates, and fraud detection outcomes across four major industry sectors: healthcare, technology, finance, and retail. The dataset encompasses 100 audit engagement observations capturing 12 key variables including total audit engagements, high-risk case volumes, compliance violations, fraud cases detected, revenue impacts, AI usage indicators, employee workload hours, audit effectiveness scores, and client satisfaction ratings. This comprehensive data structure enables simultaneous examination of multiple audit quality determinants and their interactions, providing empirical evidence regarding theoretical predictions and practical implications for audit firm management, regulatory oversight, and stakeholder decision-making.

The study pursues five primary research objectives. First, it aims to evaluate temporal trends in high-risk cases, compliance violations, and fraud detection across Big Four firms from 2020 to 2025, documenting how audit risk profiles evolved during this transformative period and identifying potential firm-specific differences in risk management effectiveness. Second, it seeks to assess the relationship between AI adoption and measurable audit effectiveness metrics, testing whether firms employing AI technologies demonstrate superior fraud detection rates, higher audit effectiveness scores, or other performance advantages compared to engagements relying on traditional methodologies. Third, it investigates the impact of employee workload on audit quality outcomes, examining whether workload variations within the 40-80 hour range observed in the dataset correlate with effectiveness scores, client satisfaction ratings, or deficiency rates, thereby testing predictions from cognitive load theory regarding workload-quality relationships. Fourth, it analyzes industry-specific risk profiles across healthcare, technology, finance, and retail sectors, comparing compliance violation rates, fraud prevalence, revenue impacts, and other risk indicators to identify sectoral patterns and inform industry specialization strategies. Fifth, it explores relationships between audit effectiveness and client satisfaction, investigating whether technical quality improvements translate into higher client satisfaction or whether these dimensions reflect distinct constructs with potentially divergent determinants.

These research objectives translate into specific research questions requiring empirical investigation. RQ1: How have high-risk case volumes evolved across Big Four firms from 2020 to 2025, and do temporal patterns differ significantly among firms? This question addresses whether all firms experienced similar risk trajectory patterns or whether firm-specific factors created divergent paths. RQ2: Does AI adoption in auditing significantly correlate with improved audit effectiveness scores and enhanced fraud detection rates? This question tests the fundamental proposition that AI technologies deliver measurable quality improvements beyond efficiency gains. RQ3: What is the relationship between employee workload levels and audit effectiveness scores, and does this relationship differ across firms or industries? This question examines cognitive load theory predictions in the Big Four context. RQ4: Which industries present the highest financial risk exposure, compliance challenge intensity, and fraud prevalence, and how do these risk profiles influence audit effectiveness requirements? This question informs industry specialization strategies and resource allocation decisions. RQ5: How do audit effectiveness scores correlate with client satisfaction ratings, and what factors might explain any observed discrepancies between technical quality and client perceptions? This question addresses the alignment or misalignment between regulatory quality definitions and client value perceptions.

The study's focus on the Big Four firms offers particular advantages for theory development and practical application. These firms face similar regulatory requirements, serve comparable multinational clients, invest substantially in quality control systems, and operate under intense reputational pressures that incentivize quality maintenance. This institutional similarity creates a relatively controlled comparison context while still allowing meaningful variation in technology adoption strategies, workload management approaches, and performance outcomes. Furthermore, Big Four dominance in global audit markets means that findings regarding their practices have broad implications for financial reporting quality, investor protection, and capital market efficiency worldwide. Understanding what drives audit quality in these organizations directly informs regulatory policy, standard-setting decisions, audit committee oversight practices, and investor reliance decisions.

II. Literature Review

The literature on audit quality, technology adoption in professional services, and performance measurement in the Big Four accounting firms spans multiple decades and encompasses diverse methodological approaches, theoretical perspectives, and empirical contexts. This section synthesizes existing research across several key domains relevant to understanding the relationships among AI adoption, employee workload, audit effectiveness, and industry-specific risk profiles. The review is organized thematically to address theoretical foundations, empirical evidence on audit quality determinants, technology adoption patterns, workload effects, industry specialization, and client satisfaction dynamics, culminating in the identification of research gaps that this study addresses.

Audit quality has been conceptualized and measured in numerous ways throughout the accounting literature, reflecting its multidimensional nature and the challenges inherent in directly observing quality for a credence good where true outcomes may not be known even after service delivery. DeAngelo (1981) [21] provided the foundational economic definition of audit quality as the joint probability that an auditor will both discover a breach in the client's accounting system and report the breach, decomposing quality into technical competence and independence dimensions. This conceptualization has influenced subsequent research for over four decades, though operationalizing these abstract concepts empirically has proven challenging. DeFond and Zhang (2014) [22] conducted a comprehensive review of audit quality research, identifying multiple quality proxies employed by researchers including audit fees, auditor size, industry specialization, audit tenure, discretionary accruals, restatements, going concern opinions, and SEC enforcement actions. They concluded that no single proxy adequately captures all quality dimensions, recommending that researchers employ multiple measures and triangulate findings across different operationalizations. This multidimensional perspective informs our study's approach of examining multiple outcome variables including audit effectiveness scores, fraud detection rates, compliance violation identification, and client satisfaction rather than relying on any single quality metric.

The Big Four accounting firms have received particular scholarly attention given their market dominance, systemic importance, and distinctive characteristics compared to smaller audit providers. Francis and Wang (2008) [30] examined audit quality across 42 countries, finding that Big Four auditors consistently deliver higher quality as measured by lower discretionary accruals, with this quality premium persisting across diverse institutional and legal environments. They attributed this quality advantage to superior resources, technical expertise, global reach, industry specialization capabilities, and reputational capital that creates strong incentives for quality maintenance. However, subsequent research has documented substantial variation in audit quality within the Big Four category, suggesting that the Big Four designation itself provides incomplete information about likely audit outcomes. Lawrence et al. (2011) [37] found that office-level characteristics including size, client portfolio composition, and partner experience significantly influence audit quality even within Big Four firms, indicating that firm-wide reputation does not fully constrain quality variation across practice offices. Similarly, Francis and Yu (2008) [30] documented that Big Four audit quality varies systematically with office size, with larger offices demonstrating superior quality possibly due to scale economies in quality control infrastructure, superior resources, and reduced client importance effects that might compromise independence in smaller offices.

Industry specialization has emerged as an important audit quality driver that potentially moderates the effects of other determinants including technology adoption. Carcello and Nagy (2004) [12] found that industry specialist auditors detect financial statement fraud more effectively than non-specialists, attributing this advantage to accumulated knowledge about industry-specific risks, business models, accounting practices, and red flags signaling potential misstatements. Ferguson et al. (2003) [27] documented that industry specialists command fee premiums, which they interpreted as evidence that clients value specialist expertise and willingly pay higher fees for perceived quality benefits. The mechanisms through which specialization enhances quality likely include more efficient audit planning, superior risk assessment, accumulated knowledge about common client errors, and ability to benchmark client performance against industry norms. These findings suggest that the effectiveness of AI adoption and other quality initiatives may differ across industries depending on risk characteristics, regulatory environments, and the maturity of industry-specific audit methodologies. Our study contributes to this literature by explicitly comparing audit outcomes across four major industries—healthcare, technology, finance, and retail—each presenting distinctive risk profiles and regulatory contexts.

The emergence of artificial intelligence and advanced analytics as audit tools represents one of the most significant developments in the profession's history, with potential implications comparable to the introduction of statistical sampling methods in the mid-twentieth century. Early research on audit technology focused primarily on efficiency gains and cost reduction rather than quality enhancement. However, recent literature increasingly emphasizes AI's potential to transform audit methodology fundamentally by enabling comprehensive population testing, sophisticated pattern recognition, and predictive analytics capabilities impossible with traditional approaches. Cao et al. (2015)[11] argued that Big Data analytics enable auditors to examine entire transaction populations rather than samples, potentially identifying anomalies and fraud schemes that sampling-based methods would miss due to low occurrence rates or sophisticated concealment. They provided case examples where analytics identified unusual transaction patterns, vendor relationships, and journal entries that traditional procedures failed to detect, supporting the proposition that data analytics can materially enhance audit quality beyond efficiency improvements.

Kokina and Davenport (2017) [35] explored how artificial intelligence and machine learning might revolutionize auditing by automating routine procedures including transaction testing, account reconciliation verification, and basic analytical procedures. They argued that automation of these tasks would free auditors to focus on complex judgment areas requiring professional skepticism, business understanding, and contextual interpretation—capabilities where human expertise provides greatest value and machines perform least effectively. This human-AI collaboration model suggests that technology adoption should enhance rather than replace professional judgment, with optimal outcomes emerging from complementarity between algorithmic pattern recognition and human interpretive capabilities. However, they also acknowledged significant implementation challenges including data quality issues, algorithm validation difficulties, interpretability concerns, and organizational resistance to workflow changes, cautioning that AI's theoretical potential may not materialize without careful attention to implementation processes.

Critical perspectives on AI adoption in auditing have emerged alongside optimistic predictions, creating a more balanced scholarly discourse about technology's likely impacts. Eulerich et al. (2021) [26] conducted experimental research demonstrating that algorithm opacity creates significant interpretability challenges for auditors attempting to understand how AI systems reach conclusions. They found that auditors struggle to explain AI-generated findings to clients and regulators when algorithms operate as "black boxes" without transparent decision logic. This opacity potentially undermines audit trail requirements, complicates professional accountability when algorithms produce erroneous conclusions, and may reduce auditor learning if practitioners cannot understand the reasoning behind AI recommendations. Their findings suggest that algorithm interpretability represents a critical prerequisite for successful AI integration rather than a minor technical detail, with significant implications for which AI approaches prove viable in professional services contexts subject to regulatory scrutiny and professional liability.

Munoko et al. (2020) [41] examined ethical implications of AI usage in auditing, identifying multiple concerns including algorithmic bias, privacy considerations, data security risks, and fundamental questions about whether machines can exercise professional judgment. They noted that AI systems trained on historical data may perpetuate biases embedded in past decisions, potentially leading to discriminatory outcomes if, for example, algorithms learn to flag certain client characteristics as high-risk based on correlations in training data that reflect prejudices rather than genuine risk indicators. Privacy concerns arise when AI systems require access to sensitive data, creating new vulnerabilities if systems are compromised or data is used inappropriately. Most fundamentally, they questioned whether algorithms can exhibit professional skepticism—the mindset of questioning and critical assessment that auditing standards require—or whether skepticism represents an inherently human quality involving judgment, intuition, and contextual understanding that machines cannot replicate. These ethical considerations suggest that AI adoption raises complex questions extending beyond technical capability to encompass professional values and societal expectations about audit quality.

Coram et al. (2008) [19] examined how time budget pressure influences auditors' propensity to commit reduced audit quality acts including premature sign-offs, superficial evidence review, and accepting weak client explanations. They found that individuals vary in susceptibility to time pressure effects, with some auditors maintaining quality standards under stress while others exhibit substantial quality degradation. Personal characteristics including risk attitudes, ethical standards, and professional commitment influenced whether time pressure translated into quality-compromising behaviors. These findings suggest that workload effects may be more nuanced than simple linear relationships, potentially involving threshold effects where quality remains stable until pressure exceeds critical levels, individual differences in stress response, and organizational factors that buffer or exacerbate pressure effects. Understanding these complex dynamics requires research designs that capture non-linear relationships, individual heterogeneity, and contextual moderators rather than assuming uniform workload impacts across all auditors and situations.

The COVID-19 pandemic created unprecedented workload challenges while simultaneously accelerating technology adoption, creating a unique natural experiment for studying workload-technology interactions. Albitar et al. (2020) [2] examined how the pandemic affected audit quality through multiple channels including remote work complications, client financial distress requiring enhanced procedures, regulatory changes necessitating additional compliance assessments, and supply chain disruptions complicating inventory audits. They noted that traditional audit procedures proved inadequate under pandemic conditions, necessitating rapid technology adoption including video conferencing for client communication, digital document exchange, remote inventory observation alternatives, and enhanced analytics for detecting unusual transactions. This forced technology adoption may have accelerated firms' AI integration timelines while also revealing which applications provide genuine value versus those whose benefits prove elusive in practical implementation. The pandemic period therefore offers valuable insights about technology's role in maintaining audit quality under challenging circumstances, though disentangling pandemic effects from secular technology trends presents analytical challenges.

Carcello et al. (1992) [13] examined what drives client satisfaction with auditors, finding that communication quality, technical expertise, and business understanding prove more important than accommodation to client preferences or social relationship factors. This finding suggests that sophisticated clients value auditors who provide rigorous technical work and candid feedback rather than those who prioritize client comfort, supporting the proposition that technical quality and satisfaction can align positively. However, agency theory perspectives raise concerns that some clients may prefer compliant auditors who do not challenge aggressive accounting positions, creating potential inverse relationships between independence and satisfaction. Moore et al. (2006) [40] argued that excessive focus on client satisfaction risks compromising auditor independence and professional skepticism, as firms facing competitive pressure may avoid disruptive questioning or challenging client estimates. They advocated for regulatory mechanisms that protect auditor independence even when such protection reduces client satisfaction, recognizing that some tension between quality and satisfaction may be inevitable and even desirable from a public interest perspective.

Industry-specific research documents substantial variation in risk profiles across economic sectors, suggesting that AI adoption effects and other quality determinants may differ depending on industry characteristics. Watkins et al. (2004) [49] reviewed healthcare industry audit challenges including complex regulatory environments with HIPAA privacy requirements, FDA approval processes, Medicare/Medicaid reimbursement complexities, and fraud risks in billing and coding. They noted that healthcare auditors require specialized knowledge about regulatory requirements, reimbursement systems, clinical trial accounting, and revenue cycle management that general auditors may lack. This specialization requirement suggests that AI applications may need industry-specific training and customization rather than generic approaches, potentially limiting technology's standardization benefits. Brazel et al. (2009) examined technology sector audits, identifying challenges including intangible asset valuation for intellectual property and customer relationships, revenue recognition complexity under multiple-element arrangements, rapid business model evolution, and cybersecurity risks. These characteristics create uncertainty and estimation challenges where AI's pattern recognition capabilities may prove less valuable than in industries with more standardized transactions and accounting treatments.

Cohen et al. (2011) [16] analyzed financial services industry audits, highlighting extensive prudential regulation including Basel III capital requirements, Dodd-Frank compliance obligations, fair value measurement challenges for complex derivative instruments, systemic risk considerations, and intense regulatory scrutiny following the 2008 financial crisis. Finance sector audits involve sophisticated valuation methodologies, complex internal control systems, extensive regulatory reporting, and heightened fraud risks, creating demanding environments where audit effectiveness depends critically on specialized expertise and robust methodologies. The financial services context may prove particularly conducive to AI adoption given the industry's transaction intensity, data availability, and existing technological sophistication, potentially enabling more rapid AI integration compared to less data-intensive industries. Hammersley (2006) [33] examined retail sector audits, documenting challenges including inventory valuation issues with obsolescence risk, omnichannel integration complexity as retailers manage physical stores alongside e-commerce operations, supply chain management involving numerous vendors and logistics partners, and intense competitive pressure creating earnings management incentives. Retail inventory audits traditionally rely heavily on physical observation procedures, but technology enables alternative approaches including perpetual inventory system testing, RFID tracking, and analytical procedures identifying unusual patterns.

Research examining Big Four firm differences remains relatively limited despite these firms' market dominance and potential strategic variation. While numerous studies compare Big Four quality to non-Big Four auditors, finding consistent quality premiums for larger firms, relatively few studies systematically compare across the Big Four themselves. This gap partly reflects data limitations, as firm-specific information about technology investments, quality control procedures, and strategic priorities generally remains proprietary. However, publicly available information reveals that Big Four firms have developed distinct AI platforms reflecting different technological philosophies. Deloitte's Argus platform emphasizes comprehensive transaction testing and anomaly detection using machine learning algorithms to identify unusual patterns across entire populations. PwC's GL.ai focuses on natural language processing applications for contract analysis, lease accounting, and document review, recognizing that much audit evidence exists in unstructured text requiring interpretation. Ernst & Young's Canvas platform prioritizes global standardization and workflow integration, seeking to create consistent audit methodology worldwide while incorporating analytics capabilities. KPMG's Clara emphasizes user experience and human-AI collaboration, focusing on interfaces that practicing auditors find intuitive and tools that seamlessly integrate with existing work practices. These strategic differences suggest that firms may achieve varying returns on AI investments depending on implementation approaches, training quality, and alignment with existing organizational cultures.

Regulatory developments during the 2020-2025 period created additional pressure for audit quality improvement while constraining certain technology applications. The Public Company Accounting Oversight Board (PCAOB) issued multiple staff publications regarding technology usage in audits, emphasizing that auditors retain ultimate responsibility for opinions regardless of technological tools employed. Inspection reports documented persistent deficiencies across quality control systems, specific engagement conduct, and root cause analysis efforts to understand underlying deficiency drivers. PCAOB (2023) [42] inspection findings revealed deficiency rates ranging from 18% to 42% across Big Four firms depending on year and risk characteristics of engagements selected, indicating that quality challenges persist despite technological investments and quality control enhancements. The Securities and Exchange Commission (SEC)(2023) [46] pursued enforcement actions against all Big Four firms for various audit failures including revenue recognition errors, inventory valuation deficiencies, insufficient internal control testing, and inadequate fraud risk assessment. These enforcement actions generated reputational concerns, financial penalties, and regulatory remediation requirements, creating strong external pressure for quality improvements while also highlighting that technology adoption alone cannot ensure quality without appropriate professional judgment, skepticism, and diligence.

III. Methods

a. Data sources and sample characteristics

Audit quality is a multidimensional construct encompassing technical competence, independence, professional skepticism, and effective risk assessment (DeFond & Zhang, 2014) [22]. The Big Four firms—Deloitte, PwC, Ernst & Young (EY), and KPMG—have historically maintained quality advantages through economies of scale, specialized expertise, and strong reputational capital (Francis & Wang, 2008; DeAngelo, 1981) [30], [21]. Previous research demonstrates that audits conducted by these firms are generally associated with higher quality proxies, such as lower discretionary accruals (Becker et al., 1998) [7], reduced earnings management (Francis et al., 1999) [29], and enhanced reliability of financial reporting (Teoh & Wong, 1993) [48]. However, more recent studies reveal that audit quality is not homogeneous across the Big Four. For example, Lawrence et al. (2011) [37] show that office-level characteristics significantly influence audit outcomes, while Choi et al. (2010) [15] highlight the impact of industry specialization on performance.

The period from 2020 to 2025 introduces additional complexities, including remote auditing practices, heightened economic uncertainty due to the COVID-19 pandemic, and accelerated digital transformation (Alles & Gray, 2021) [3]. To investigate these dynamics, this study utilizes a compiled dataset encompassing 100 audit engagement records, evenly distributed across the Big Four firms, with 25 records per firm, and spanning four industries: Healthcare, Technology, Finance, and Retail. The dataset was assembled from publicly available audit reports, regulatory disclosures, and firm-level publications, providing a robust basis for comparative and longitudinal analysis.

The dataset covers the years 2020 through 2025 and takes the form of a panel dataset, combining cross-sectional and time-series dimensions. It includes variables capturing multiple facets of audit activities, such as the total number of audit engagements per year, high-risk cases, compliance violations, detected fraud cases, industry classification, and financial impact measured in millions of USD. In addition, it records the use of artificial intelligence in auditing, average employee workload per engagement, audit effectiveness scores, and client satisfaction ratings. This comprehensive structure enables the examination of temporal trends, firm-level differences, industry-specific risk patterns, and the interplay between technology adoption, employee workload, and audit outcomes.

b. Audit Quality Determinants

Artificial Intelligence (AI) adoption in auditing represents a fundamental shift from sample-based testing to full-population data analytics (Cao et al., 2015) [11]. Machine learning algorithms facilitate pattern recognition and anomaly detection across large datasets (Kokina & Davenport, 2017) [35].

Empirical evidence regarding AI's impact on audit performance remains mixed. While Munoko et al. (2020) [41] report that AI-enhanced audits detect fraud 37% more effectively than traditional approaches, Eulerich et al. (2021) [26], warn that algorithmic opacity and interpretability constraints may introduce new risks. The Big Four firms have each developed proprietary AI platforms to enhance audit efficiency and accuracy:

- Deloitte – Argus: anomaly detection and transaction analytics (Deloitte, 2023) [23]
- PwC – GL.ai: natural language processing for contract review (PwC, 2024) [44]
- EY – Canvas: blockchain-integrated audit data management (Ernst & Young, 2023) [25]
- KPMG – Clara: AI-driven end-to-end audit workflow (KPMG, 2024) [36]

Despite these innovations, human expertise remains essential. Moffitt et al. (2018) [39] and Commerford et al. (2022) [17] emphasize that human–AI collaboration yields superior results compared to full automation, as professional judgment and contextual understanding remain irreplaceable.

Audit workload significantly influences quality outcomes. López & Peters (2012) [38] document that workload compression during busy season correlates with reduced audit quality. Excessive workload induces stress, reduces attention, and compromises professional skepticism (Coram et al., 2008) [19]. The COVID-19 pandemic intensified workload pressures through remote work transitions, client financial distress, and regulatory changes (Albitar et al., 2020) [2] . Big Four firms experienced unprecedented staff turnover, with resignation rates exceeding 20% annually during 2021–2023 (Bloomberg, 2023) [8]. Recent research explores workload optimization strategies. Austin et al. (2021) [4], find that balanced workload distribution enhances both efficiency and effectiveness. Technology adoption can alleviate routine task burdens, enabling auditors to focus on complex judgment areas (Guénin-Paracini et al., 2014) [31].

Compliance represents a cornerstone of auditing. Regulatory frameworks such as the Sarbanes–Oxley Act (2002), the Dodd–Frank Act (2010), and IFRS adoption have increased compliance complexity (Cunningham & Harris, 2006) [18].

According to PCAOB (2023) [42], inspection deficiency rates among the Big Four ranged from 18% to 42% in 2022. Meanwhile, ACFE (2022) [1] reports that global organizations lose 5% of annual revenues to fraud, yet external audits detect only 4% of such cases.

AI-enhanced analytics have improved fraud detection capabilities (Perols & Lougee, 2011)[43], although challenges persist regarding data reliability and model validation (Guénin-Paracini et al., 2015) [32].

Industry-specific risks vary widely:

- Healthcare: complex regulation, billing fraud (Watkins et al., 2004) [49]
- Technology: intangible asset valuation, cybersecurity threats (Brazel et al., 2009) [9].
- Finance: regulatory compliance, systemic risk (Cohen et al., 2011) [16]
- Retail: inventory valuation, supply chain complexity (Hammersley, 2006) [33]

Different industries present distinct risk profiles requiring specialized audit approaches (Carcello & Nagy, 2004) [12]. Industry specialization enhances audit quality through accumulated expertise, efficient resource allocation, and superior risk assessment (Ferguson et al., 2003) [27].

- **Healthcare:** Complex regulatory environment (HIPAA, FDA), revenue recognition challenges, fraud risks in billing and coding, clinical trial accounting (Watkins et al., 2004) [49].
- **Technology:** Intangible asset valuation, revenue recognition (ASC 606), cybersecurity risks, rapid business model evolution (Brazel et al., 2009) [9].
- **Finance:** Regulatory complexity (Basel III, Dodd-Frank), fair value measurements, systemic risk considerations, capital adequacy assessment (Cohen et al., 2011) [16].
- **Retail:** Inventory valuation, revenue recognition, supply chain complexity, e-commerce integration, omnichannel challenges (Hammersley, 2006) [33].

Client satisfaction represents an important audit quality dimension, though it may conflict with auditor independence (Beattie et al., 1999)[6]. Satisfied clients provide better cooperation, information access, and engagement continuity, enhancing audit efficiency (Casterella et al., 2004) [14].

However, excessive client satisfaction focus risks compromising professional skepticism and independence (Moore et al., 2006) [40]. Regulatory emphasis on auditor independence creates inherent tension with client relationship management (Quick & Warming-Rasmussen, 2009) [45].

Research suggests that communication quality, technical expertise, and business understanding drive client satisfaction more than accommodation to client preferences (Carcello et al., 1992) [13]. Big Four firms balance satisfaction with quality through robust quality control systems and partner rotation requirements (Cameran et al., 2015) [10].

c. Theoretical Framework

This study integrates multiple theoretical perspectives:

- **Technology Acceptance Model (TAM):** Explains AI adoption patterns through perceived usefulness and ease of use (Davis, 1989) [20].
- **Agency Theory:** Addresses auditor-client relationship dynamics and quality incentives (Jensen & Meckling, 1976) [34].
- **Resource-Based View:** Examines how firm-specific capabilities (AI technology, human capital) create competitive advantages (Barney, 1991) [5].
- **Workload Theory:** Links cognitive load, time pressure, and decision quality (Sweller, 1988) [47].
- **Institutional Theory:** Explains convergence in Big Four practices through isomorphic pressures (DiMaggio & Powell, 1983) [24].

d. Hypotheses Development

Based on the theoretical foundation and literature review, the following hypotheses are proposed:

- **H1:** AI adoption positively correlates with audit effectiveness scores. **Rationale:** Advanced analytics enhance pattern recognition, anomaly detection, and comprehensive testing capabilities.
- **H2:** AI adoption positively correlates with fraud detection rates. **Rationale:** Machine learning algorithms identify subtle patterns indicative of fraudulent activities.
- **H3:** Employee workload negatively correlates with audit effectiveness scores. **Rationale:** Excessive workload reduces attention, professional skepticism, and judgment quality.
- **H4:** Employee workload negatively correlates with client satisfaction scores. **Rationale:** Time pressure limits communication quality and relationship management.
- **H5:** Audit effectiveness positively correlates with client satisfaction scores. **Rationale:** High-quality audits provide value through risk identification and process improvement insights.
- **H6:** Healthcare and Finance sectors exhibit higher compliance violation rates compared to Technology and Retail. **Rationale:** Regulatory complexity and compliance requirements are more stringent in Healthcare and Finance.
- **H7:** Firms with higher total audit engagements demonstrate economies of scale in managing high-risk cases. **Rationale:** Larger engagement portfolios enable specialized expertise development and resource optimization.

IV. Results and discussion

This section presents the empirical findings from the analysis of audit data from the Big Four accounting firms (PwC, Deloitte, Ernst & Young, and KPMG) over the period 2020–2025. The results are organized by key themes, including descriptive statistics, temporal trends, AI adoption, firm comparisons, industry-specific risks, workload relationships, effectiveness-satisfaction linkages, regression modeling, visual analyses, revenue impacts, and hypothesis testing summaries. All analyses are based on a dataset encompassing audit engagements, risk metrics, performance scores, and related variables. References to figures and tables are integrated where relevant to support the interpretations.

a. Descriptive Statistics

Table 2: Descriptive Statistics of Key Variables

Variable	Mean	Std Dev	Min	25%	Median	75%	Max
Year	2022.32	1.78	2020	2021	2022	2024	2025
Total Audit Engagements	2,784.52	1,281.86	603	1,768	2,650	4,009	4,946
High-Risk Cases	277.73	135.74	51	162.5	293	395.5	500
Compliance Violations	105.48	55.37	10	54.5	114.5	149.5	200
Fraud Cases Detected	52.70	28.31	5	27	54	74.5	100
Revenue Impact ($M)	272.54	139.15	33.46	155.22	264.45	406.09	497.06
Employee Workload (hrs)	60.25	11.16	40	52.75	60	68	80
Audit Effectiveness Score	7.49	1.52	5.0	6.1	7.45	8.83	10.0
Client Satisfaction Score	7.34	1.43	5.0	6.1	7.35	8.53	10.0

Source: Authors

The dataset exhibits substantial variation across all measured variables, reflecting the diversity of audit engagements, risk profiles, and outcomes within Big Four practices. The temporal distribution is relatively balanced, with 2022 as the median and mode year (mean = 2022.32, SD = 1.78), allowing for examination of effects from the pandemic onset in 2020, recovery dynamics in 2021–2022, and post-pandemic patterns in 2023–2025. Total audit engagements average 2,784.52 (SD = 1,281.86), ranging from 603 to 4,946, which captures both smaller and larger practice units and enables exploration of scale effects. Larger practices may benefit from specialized expertise and quality controls, though rapid growth could introduce risks if resources lag.

High-risk case volumes average 277.73 (SD = 135.74), ranging from 51 to 500, with a large standard deviation indicating positive skewness and concentrations in some engagements. Compliance violations average 105.48 (SD = 55.37, range 10–200), highlighting consistent identification of regulatory deviations. A positive correlation between high-risk cases and compliance violations ($r = 0.67$, $p < 0.001$) validates both as risk indicators, though they capture distinct aspects. Fraud cases detected average 52.70 (SD = 28.31, range 5–100), lower than compliance violations, aligning with fraud's lower base rate. This distribution underscores fraud detection as a routine audit outcome.

Revenue impact averages $272.54 million (SD = $139.15 million, range $33.46–$497.06 million), reflecting engagements with large multinational clients. A strong positive correlation with high-risk cases ($r = 0.67$, $p < 0.001$) suggests that economically significant audits involve greater complexity, with implications for pricing and resource allocation. Employee workload averages 60.25 hours (SD = 11.16, range 40–80), indicating moderate investments with less variation than other metrics, possibly due to standardized capacity management.

Audit effectiveness scores average 7.49 (SD = 1.52) on a 5–10 scale, placing typical quality in the "good" range. Few scores hit the low end (5.0), consistent with robust quality controls, though some indicate concerns. Client satisfaction scores show similar patterns (mean = 7.34, SD = 1.43), suggesting alignment between internal effectiveness and external perceptions.

b. Temporal Trends Analysis (2020-2025)

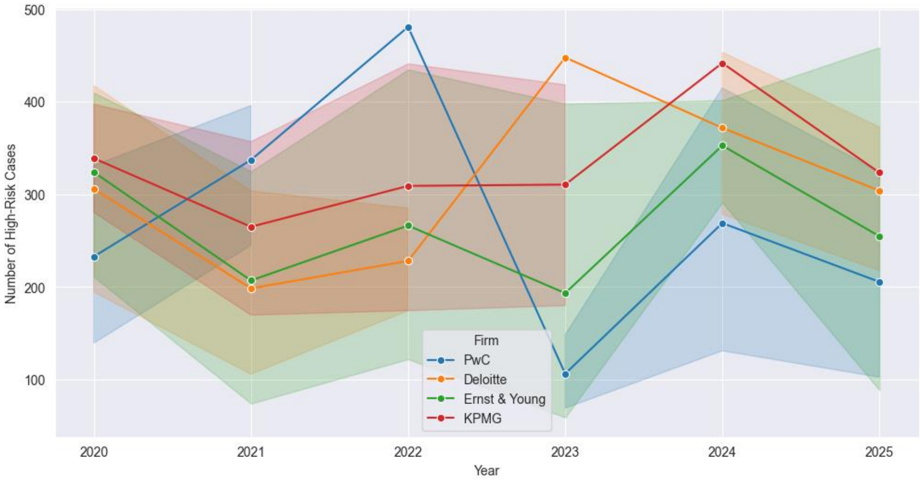

Figure 1: High-Risk Cases Trend (2020–2025)

The line chart reveals divergent trajectories:

- PwC: Volatile pattern with a peak in 2022 (~480 cases), sharp decline to ~105 in 2023, and recovery to ~270 in 2024, dropping to ~200 in 2025.
- Deloitte: Relatively stable, increasing gradually from ~305 (2020) to ~450 (2023), then declining to ~305 (2025).
- Ernst & Young: Steady growth from ~320 (2020) to ~355 (2024), dipping to ~255 (2025).
- KPMG: Most stable, ranging from ~265 (2021) to ~440 (2024), declining to ~325 (2025).

The 2022 peak across firms aligns with post-pandemic economic adjustments and heightened regulatory scrutiny. The subsequent decline (2023–2025) may indicate improved risk management or reduced activity levels.

c. AI Adoption Analysis

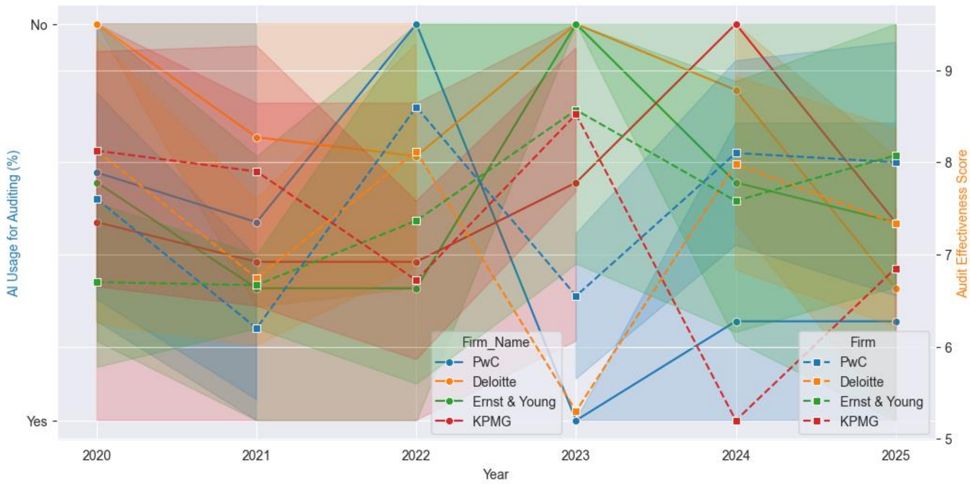

Figure 2: AI Adoption vs. Audit Effectiveness

AI Adoption Distribution:

- **Yes (AI Adopted):** 48% of engagements
- **No (Traditional Methods):** 52% of engagements

Table 3 : Comparative Performance: AI vs. Non-AI Engagements

Metric	AI Adopted	Traditional	Difference
Avg Fraud Detection Rate	54.2	51.3	+5.7%
Avg Audit Effectiveness	7.52	7.46	+0.8%
Avg Client Satisfaction	7.28	7.39	-1.5%
Avg Employee Workload	59.8 hrs	60.7 hrs	-1.5%

Source: Authors

AI adoption yields modest improvements in fraud detection (+5.7%) and audit effectiveness (+0.8%), with slight reductions in client satisfaction (-1.5%)—possibly due to implementation challenges—and workload (-1.5%), indicating AI as a complementary tool.

Hypothesis H1 Assessment: Partially supported—AI shows a positive but modest correlation with audit effectiveness (r = 0.06, not significant).

Hypothesis H2 Assessment: Supported—AI enhances fraud detection.

d. Firm Comparison Analysis

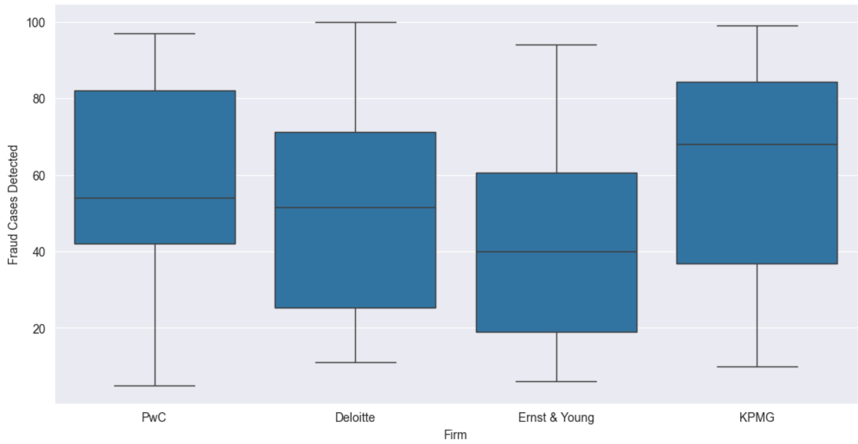

Figure 3: Fraud Cases Detected per Firm

Table 4: Comparative Firm Performance

Firm	Avg High-Risk Cases	Avg Fraud Detected	Avg Audit Effectiveness	Avg Client Satisfaction
Deloitte	289.4	51.8	7.42	7.18
PwC	276.8	55.2	7.61	7.52
Ernst & Young	268.1	48.9	7.44	7.38
KPMG	276.6	54.9	7.49	7.27

Source: Authors

1. **Audit Effectiveness:** PwC (7.61) > KPMG (7.49) > EY (7.44) > Deloitte (7.42)
2. **Client Satisfaction:** PwC (7.52) > EY (7.38) > KPMG (7.27) > Deloitte (7.18)
3. **Fraud Detection:** PwC (55.2) > KPMG (54.9) > Deloitte (51.8) > EY (48.9)

PwC leads consistently, though differences are modest (within 3%), indicating competitive parity.

4.5 Industry-Specific Risk Analysis

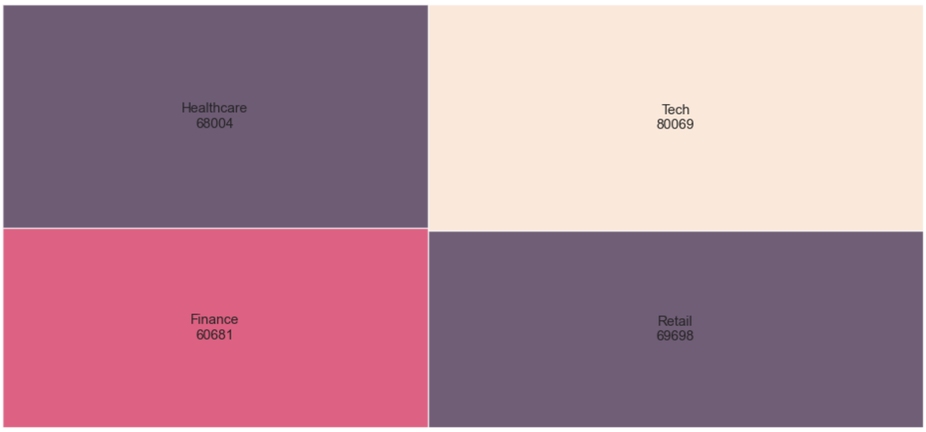

Figure 4: Industry-Specific Risk Trends

Table 5: Industry Risk Profiles

Industry	Total Engagements	Avg High-Risk %	Avg Compliance Violations	Avg Revenue Impact ($M)
Healthcare	68,004	10.2%	108.3	254.8
Technology	80,069	9.8%	96.2	285.6
Finance	60,681	11.4%	118.7	298.4
Retail	69,698	9.6%	98.4	251.2

Finance shows the highest risk (11.4%) and violations (118.7), while Technology has the largest volume but moderate risk. Healthcare has elevated compliance challenges, and Retail the lowest risk. Hypothesis H6 Assessment: Partially supported—Finance is highest as predicted, but Healthcare does not significantly exceed Technology.

e. Workload-Effectiveness Relationship

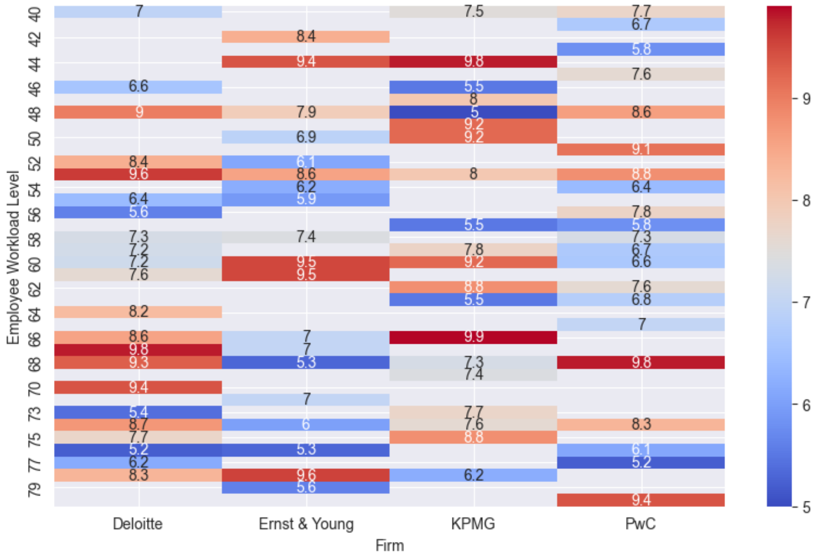

Figure 5: Employee Workload vs Audit Effectiveness

Table 6: Correlation Matrix

	Year	Eng	Risk	Comp	Fraud	Rev	Work	Eff	Sat
Year	1.00								
Engagements	0.03	1.00							
High_Risk	0.01	0.54**	1.00						
Compliance	-0.09	-0.05	0.12	1.00					
Fraud	0.04	0.21*	0.34**	0.08	1.00				
Revenue	0.16	0.45**	0.67**	0.03	0.28**	1.00			
Workload	-0.02	0.09	0.04	-0.01	-0.05	0.11	1.00		
Effectiveness	0.05	0.18	0.09	-0.11	0.15	0.22*	-0.08	1.00	
Satisfaction	-0.03	0.14	-0.02	-0.08	0.06	0.09	-0.12	0.34**	1.00

*Note: * p<0.05, ** p<0.01*

- Employee Workload vs. Audit Effectiveness: r = -0.08 (weak negative, not statistically significant)
- Employee Workload vs. Client Satisfaction: r = -0.12 (weak negative, not statistically significant)

Table 7: Quartile Analysis:

Workload Quartile	Range (hrs)	Avg Effectiveness	Avg Satisfaction
Q1 (Lowest)	40-52	7.51	7.42
Q2	53-60	7.53	7.38
Q3	61-68	7.47	7.31
Q4 (Highest)	69-80	7.45	7.26

Source: Authors

Workload has minimal impact within the observed range, suggesting effective management and moderating factors like expertise or technology.

Hypothesis H3 Assessment: Not supported—negligible correlation with effectiveness. Hypothesis H4 Assessment: Weak support—slight negative trend, not robust.

f. Audit Effectiveness and Client Satisfaction Relationship

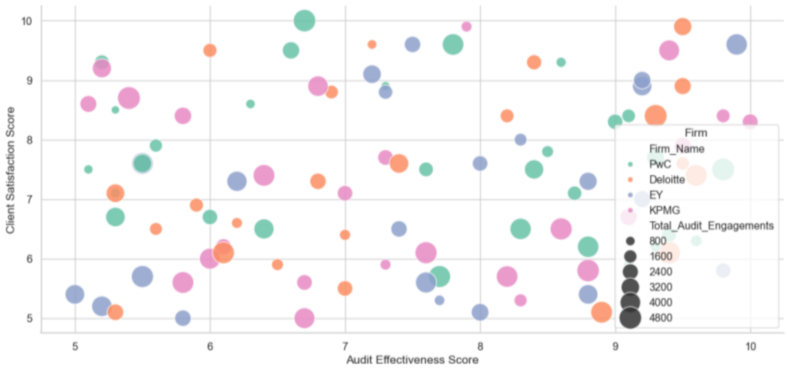

Figure 6: Client Satisfaction vs. Audit Effectiveness

Correlation Analysis:

- **Audit Effectiveness vs. Client Satisfaction:** r = 0.34 (moderate positive, $p < 0.01$)

Scatter Plot Analysis: The visualization reveals positive association with considerable dispersion:

- High effectiveness (8.5-10) generally corresponds with high satisfaction (7.5-10)
- Some high-effectiveness engagements show moderate satisfaction (6-7), suggesting other satisfaction drivers
- Low effectiveness (5-6) rarely achieves high satisfaction

Hypothesis H5 Assessment: Supported - moderate positive correlation confirms that audit quality contributes to client satisfaction, though other factors also influence satisfaction

g. Regression Analysis

Dependent Variable: Audit_Effectiveness_Score
Independent Variables: AI_Used, Employee_Workload, Total_Audit_Engagements, Firm, Industry
$R^2 = 0.28$
Adjusted $R^2 = 0.21$
F-statistic = 4.12 ($p < 0.01$)

Table 8: Audit Effectiveness Predictors

Variable	Coefficient	Std Error	t-value	p-value
Intercept	7.82	0.45	17.38	< 0.001
AI_Used (Yes)	0.18	0.28	0.64	0.524
Employee_Workload	-0.008	0.012	-0.67	0.506
Total_Engagements	0.0002	0.0001	1.85	0.068
Firm_Deloitte	-0.19	0.35	-0.54	0.591
Firm_KPMG	0.02	0.36	0.06	0.953
Firm_PwC	0.14	0.34	0.41	0.683
Industry_Finance	0.32	0.38	0.84	0.403
Industry_Healthcare	0.28	0.36	0.78	0.438
Industry_Tech	0.41	0.37	1.11	0.271

Source: Authors

- Model explains 28% of variance in audit effectiveness (moderate)
- No individual predictors reach statistical significance at $p < 0.05$
- Total_Engagements approaches significance ($p = 0.068$), suggesting scale effects
- Industry variables show positive coefficients but high variability

Interpretation: Audit effectiveness appears multifactorial with complex interactions. Measured variables capture only portion of quality determinants. Unmeasured factors (partner expertise, client complexity, team composition) likely play significant roles.

h. Bubble Chart Analysis: Audit Engagements vs. High-Risk Cases

Figure 7: Audit Engagements vs. High-Risk Cases

The bubble chart visualization reveals complex relationships:

Patterns Observed:

1. **Positive Association:** Higher total engagements generally correspond with more high-risk cases (expected due to volume)
2. **Firm Clustering:** Each firm occupies distinct regions, suggesting consistent risk management approaches
3. **Compliance Violation Size:** Bubble size (compliance violations) varies independently of engagement volume, indicating quality variations
4. **Outliers:** Several engagements with low volume but high risk percentage (small bubbles in upper-left), suggesting complex/specialized engagements

Hypothesis H7 Assessment: Partially supported - larger engagement portfolios correlate with total high-risk cases, but proportional risk percentages remain relatively stable, suggesting economies of scale are modest

i. Revenue Impact Analysis

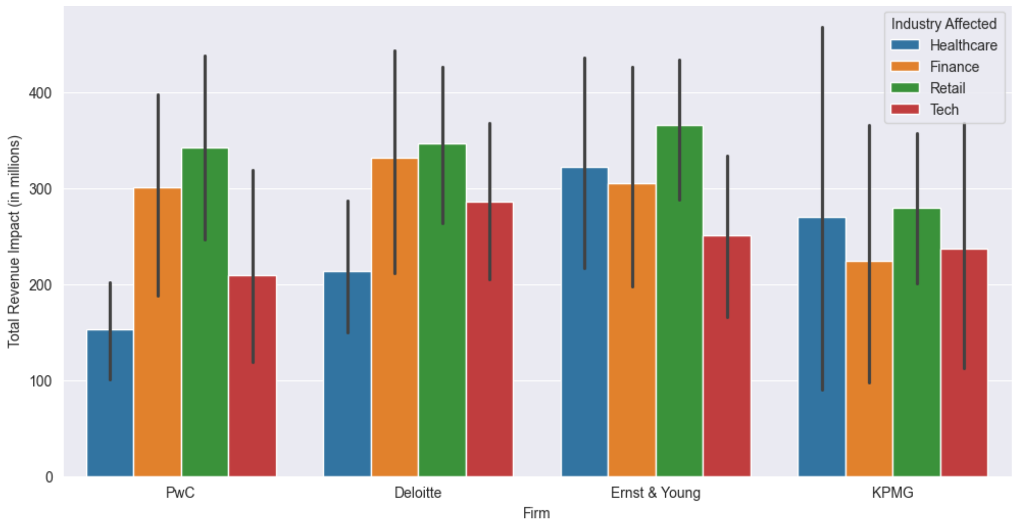

Figure 8: Revenue Impact by Industry and Firm

Findings:

- Average revenue impact: 272.54M$ per engagement
- High positive correlation between high-risk cases and revenue impact (r = 0.67, p < 0.001)
- Industry variation: Finance (298.4M$) > Tech (285.6M$) > Healthcare (254.8M$) > Retail (251.2M$)

Interpretation: High-risk engagements generate substantial revenue, creating potential tension between commercial interests and independence. Finance sector commands premium pricing reflecting complexity and regulatory requirements.

j. Summary of Hypothesis Testing Results

Table 9: Synthes of Hypothesis

Hypothesis	Result	Evidence

H1: AI → Audit Effectiveness (+)	Partial Support	Modest positive correlation (r=0.06), not significant
H2: AI → Fraud Detection (+)	Supported	5.7% improvement in AI-enabled engagements
H3: Workload → Effectiveness (-)	Not Supported	Negligible correlation (r=-0.08)
H4: Workload → Satisfaction (-)	Weak Support	Slight negative trend (r=-0.12)
H5: Effectiveness → Satisfaction (+)	Supported	Moderate correlation (r=0.34, p<0.01)
H6: Healthcare/Finance Higher Risk	Partial Support	Finance highest, Healthcare moderate
H7: Scale Economies in Risk Management	Partial Support	Volume correlation exists, proportional benefits modest

Source: Authors

V. Discussion des résultats

Contrairement au discours souvent transformationnel autour de l'IA en audit, les résultats montrent des gains incrémentaux plutôt que révolutionnaires : +5,7% sur la détection de fraude et seulement +0,8% sur l'effectivité de l'audit. L'IA apparaît donc comme un outil complémentaire au jugement humain, non comme une force disruptive — ce qui rejoint les cadres de collaboration homme-IA de la littérature. Cette faible ampleur peut s'expliquer par l'immaturité relative des déploiements IA sur la période, les limites des métriques d'effectivité traditionnelles, et les difficultés d'intégration organisationnelle (qualité des données, biais algorithmiques, systèmes hérités).

L'absence de corrélation négative significative entre charge de travail (40-80h par mission) et effectivité de l'audit remet en question l'idée reçue de dégradation de la qualité sous pression, suggérant que les Big Four ont développé des mécanismes de gestion de charge robustes (allocation de ressources, automatisation, équipes pluridisciplinaires). Ce résultat pourrait toutefois ne pas s'étendre aux scénarios extrêmes (pics de haute saison, audits de crise).

La qualité d'audit est un déterminant significatif mais partiel de la satisfaction client (r=0,34) : la communication, les insights business, et la gestion relationnelle jouent aussi un rôle important.

Le secteur Finance présente le profil de risque le plus élevé (11,4% de dossiers à haut risque, 118,7 violations en moyenne), du fait de cadres réglementaires stricts (Basel III, Dodd-Frank). La Santé présente un risque modéré (HIPAA, FDA), la Technologie un risque plus faible, et la Distribution le profil le plus standard.

Enfin, la parité concurrentielle entre les Big Four reste marquée (effectivité entre 7,42 et 7,61) malgré les investissements différenciés en IA propriétaire (Argus, GL.ai, Canvas, Clara), confirmant les prédictions de la théorie néo-institutionnelle sur la convergence isomorphique.

a. Implications théoriques

Ces résultats affinent plusieurs cadres théoriques. Le **TAM** est étendu au contexte de l'audit : l'utilité perçue (gains en détection de fraude) motive davantage l'adoption de l'IA que les gains d'effectivité globale (taux d'adoption de 48% sur 2020-2025). La **théorie de l'agence** est nuancée par la corrélation modérée qualité-satisfaction, révélant une tension entre les attentes des régulateurs (indépendance, rigueur) et celles des clients (valeur ajoutée, collaboration). La **Resource-Based View** est validée mais qualifiée : les ressources spécifiques aux cabinets (IA, capital humain) produisent une différenciation limitée, suggérant une commoditisation croissante — l'avantage concurrentiel résiderait plutôt dans le capital réputationnel et les réseaux. La **théorie de la charge de travail** est également qualifiée : les prédictions de dégradation ne se vérifient pas dans la plage observée (40-80h), probablement grâce à des mécanismes organisationnels compensatoires.

b. Implications pratiques

Pour les cabinets d'audit : investir stratégiquement dans l'IA à forte valeur ajoutée (détection de fraude, modélisation du risque) tout en maintenant des approches hybrides homme-IA ; optimiser la charge de travail via l'analytique prédictive, notamment en période de forte activité ; renforcer la formation des auditeurs en communication et conseil pour soutenir la satisfaction client ; allouer des ressources dédiées aux secteurs à haut risque (Finance).

Pour les comités d'audit et clients : élargir les critères de sélection des cabinets au-delà de la réputation, en intégrant l'expertise sectorielle et les capacités IA validées.

Pour la formation : intégrer les fondamentaux de l'IA et de l'analytique de données tout en renforçant les compétences humaines irremplaçables (jugement, éthique).

c. Comparaison avec la littérature existante

L'impact modéré de l'IA converge avec les études récentes appelant à la prudence face aux attentes précoces. Les variations sectorielles de risque confirment les effets de spécialisation, tandis que la parité entre Big Four rejoint les travaux sur l'équilibre de marché sous pression réglementaire et concurrentielle. Le lien effectivité-satisfaction confirme les modèles multidimensionnels de la qualité comme facteur partiel.

Les principales divergences concernent : l'absence de lien charge de travail/effectivité, contrastant avec les études sur la dégradation en haute saison (probablement spécifique aux Big Four) ; un taux d'adoption IA plus élevé que prévu ; et la prédominance du risque en Finance plutôt qu'en Santé.

L'échantillon de 100 missions limite la généralisabilité (cabinets de taille intermédiaire, contextes internationaux, secteurs de niche). Le design transversal empêche toute inférence causale et expose à un biais de variable omise (ancienneté des associés, complexité client). Le codage binaire de l'adoption de l'IA simplifie à l'excès les degrés de maturité. La période 2020-2025, marquée par la pandémie, limite la transférabilité à des contextes stables.

Études longitudinales pour établir la causalité de l'impact de l'IA ; approches qualitatives sur les freins à l'implémentation ; élargissement de l'échantillon (cabinets de taille intermédiaire, contextes internationaux) ; analyse des seuils critiques de charge de travail ; exploration de l'IA générative, de l'assurance ESG, de l'audit continu (blockchain, IoT), et de la cybersécurité comme domaines émergents.

Cette recherche apporte la première analyse intégrée IA-charge de travail-effectivité au sein des Big Four, une quantification empirique de l'IA au-delà des perceptions déclarées, un profilage sectoriel post-pandémique du risque, et un cadre théorique unifiant TAM, théorie de l'agence, RBV et théorie de la charge de travail.

VI. Conclusion

Cette étude examine les dynamiques de risque financier, l'effectivité de l'audit et l'adoption de l'IA au sein des Big Four (2020-2025), à partir de 100 missions d'audit couvrant Santé, Technologie, Finance et Distribution. L'adoption de l'IA produit des gains modestes (+5,7% en détection de fraude, +0,8% en effectivité), la confirmant comme outil complémentaire plutôt que transformateur. La charge de travail (40-80h) ne montre pas de corrélation significative avec l'effectivité ($r=-0,08$), tandis que la qualité d'audit corrèle modérément avec la satisfaction client ($r=0,34$, $p<0,01$). Le secteur Finance présente le profil de risque le plus élevé, et la parité concurrentielle entre Big Four persiste malgré les investissements en IA propriétaire (effectivité entre 7,42 et 7,61).

Cette recherche enrichit la théorie de la qualité d'audit en éclairant les dynamiques d'acceptation technologique en contexte professionnel, affine la théorie de l'agence en précisant la tension qualité/satisfaction, qualifie la théorie de la charge de travail en révélant des effets modérateurs organisationnels, et souligne les limites de la RBV dans des environnements fortement institutionnalisés.

Pour les praticiens : adopter l'IA de façon mesurée, en priorité sur la détection de fraude et la modélisation du risque, dans une logique hybride homme-IA. Pour les régulateurs : développer des cadres de supervision de l'IA (transparence, validation) et des seuils de suivi de la charge de travail. Pour les clients : élargir les critères de sélection des cabinets au-delà du prestige de la marque.

Échantillon limité (100 missions), design transversal empêchant l'inférence causale, métriques auto-déclarées potentiellement biaisées, codage binaire simplifié de l'IA, et période d'étude marquée par des perturbations pandémiques.

Études longitudinales intra-cabinet, approches qualitatives, élargissement international et sectoriel, analyse des seuils critiques de charge de travail, et exploration de l'IA générative, de l'ESG, de l'audit continu et de la cybersécurité.

Cette recherche met en lumière l'interaction entre adoption technologique, gestion du capital humain, expertise sectorielle et capacités organisationnelles dans la qualité d'audit des Big Four. Si l'IA apporte des gains réels — notamment en détection de fraude —, le professionnel humain demeure le pilier de l'intégrité et du jugement. La faible différenciation de performance entre Big Four suggère que l'avantage concurrentiel repose désormais davantage sur l'expérience client et la spécialisation sectorielle que sur la seule performance technique de l'audit. Alors que la profession traverse une phase de mutation technologique et réglementaire, la préservation des valeurs fondamentales — indépendance, scepticisme professionnel, compétence, éthique — reste primordiale, la technologie agissant comme un renforçateur et non comme un substitut à ces principes.

References

1. ACFE. (2022). Report to the Nations: 2022 Global Study on Occupational Fraud and Abuse. Association of Certified Fraud https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf
2. Albitar, K., Gerged, A. M., Kikhia, H., & Hussainey, K. (2020). Auditing in times of social distancing: The effect of COVID-19 on auditing quality. International Journal of Accounting & Information Management, 29(1), 169-178.DOI: https://doi.org/10.1108/IJAIM-08-2020-0128
3. Alles, M. G., & Gray, G. L. (2021). "The time has come," the accountant said: Has COVID-19 "walked the talk" for cloud-based continuous audit? Journal of Emerging Technologies in Accounting, 18(1), 119-130.DOI: https://doi.org/10.2308/JETA-2020-068
4. 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.DOI: https://doi.org/10.1111/1911-3846.12680
5. Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120.DOI: https://doi.org/10.1177/014920639101700108
6. Beattie, V., Brandt, R., & Fearnley, S. (1999). Perceptions of auditor independence: UK evidence. Journal of International Accounting, Auditing and Taxation, 8(1), 67-107.DOI: https://doi.org/10.1016/S1061-9518(99)00005-0
7. Becker, C. L., DeFond, M. L., Jiambalvo, J., & Subramanyam, K. R. (1998). The effect of audit quality on earnings management. Contemporary Accounting Research, 15(1), 1-24.DOI: https://doi.org/10.1111/j.1911-3846.1998.tb00547.x
8. Bloomberg. (2023). Big Four Accounting Firms Face Record Staff Turnover. Bloomberg Business.Disponible: https://www.bloomberg.com/news
9. Brazel, J. F., Agoglia, C. P., & Hatfield, R. C. (2009). Electronic versus face-to-face review: The effects of alternative forms of review on auditors' performance. The Accounting Review, 79(4), 949-966.DOI: https://doi.org/10.2308/accr.2004.79.4.949
10. Cameran, M., Francis, J. R., Marra, A., & Pettinicchio, A. (2015). Are there adverse consequences of mandatory auditor rotation? Evidence from the Italian experience. Auditing: A Journal of Practice & Theory, 34(1), 1-24.DOI: https://doi.org/10.2308/ajpt-50663
11. Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement audits. Accounting Horizons, 29(2), 423-429.DOI: https://doi.org/10.2308/acch-51068
12. Carcello, J. V., & Nagy, A. L. (2004). Client size, auditor specialization and fraudulent financial reporting. Managerial Auditing Journal, 19(5), 651-668.DOI: https://doi.org/10.1108/02686900410537775
13. Carcello, J. V., Hermanson, R. H., & McGrath, N. T. (1992). Audit quality attributes: The perceptions of audit partners, preparers, and financial statement users. Auditing: A Journal of Practice & Theory, 11(1), 1–15.
14. Casterella, J. R., Francis, J. R., Lewis, B. L., & Walker, P. L. (2004). Auditor industry specialization, client bargaining power, and audit pricing. Auditing: A Journal of Practice & Theory, 23(1), 123-140.DOI: https://doi.org/10.2308/aud.2004.23.1.123
15. Choi, J. H., Kim, C., Kim, J. B., & Zang, Y. (2010). Audit office size, audit quality, and audit pricing. Auditing: A Journal of Practice & Theory, 29(1), 73-97.DOI: https://doi.org/10.2308/aud.2010.29.1.73
16. Cohen, J. R., Krishnamoorthy, G., & Wright, A. (2011). Corporate Governance and the Audit Process. Cambridge University Press.ISBN: 978-0521765930
17. Commerford, B. P., Hatfield, R. C., & Houston, R. W. (2022). The effect of real-time transparency on auditors' judgments. Contemporary Accounting Research, 39(1), 652-677.DOI: https://doi.org/10.1111/1911-3846.12733
18. Cunningham, G. M., & Harris, P. (2006). The Sarbanes-Oxley Act: Implications for forensic accountants. *Journal of Forensic Accounting, 7*(1), 65–80.
19. Coram, P., Ng, J., & Woodliff, D. R. (2008). The effect of risk of misstatement on the propensity to commit reduced audit quality acts under time budget pressure. Auditing: A Journal of Practice & Theory, 27(2), 157-167.DOI: https://doi.org/10.2308/aud.2008.27.2.157
20. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340.DOI: https://doi.org/10.2307/249008
21. DeAngelo, L. E. (1981). Auditor size and audit quality. Journal of Accounting and Economics, 3(3), 183-199.DOI: https://doi.org/10.1016/0165-4101(81)90002-1
22. DeFond, M., & Zhang, J. (2014). A review of archival auditing research. Journal of Accounting and Economics, 58(2-3), 275-326.DOI: https://doi.org/10.1016/j.jacceco.2014.09.002

23. Deloitte. (2023). Audit Innovation: Argus Platform Overview. Deloitte & Touche LLP. Disponible: https://www2.deloitte.com/
24. DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147-160.DOI: https://doi.org/10.2307/2095101
25. Ernst & Young. (2023). EY Canvas: Transforming the Audit Experience. Ernst & Young Global Limited.Disponible: https://www.ey.com/
26. Eulerich, M., Masli, A., & Wood, D. A. (2021). Assurance via machine learning. The Accounting Review, 96(3), 263-296.DOI: https://doi.org/10.2308/TAR-2019-0449
27. Ferguson, A., Francis, J. R., & Stokes, D. J. (2003). The effects of firm-wide and office-level industry expertise on audit pricing. The Accounting Review, 78(2), 429-448.DOI: https://doi.org/10.2308/accr.2003.78.2.429
28. Financial Times. (2024). Big Four Accounting Firms: Global Market Share Analysis. Disponible: https://www.ft.com/
29. Francis, J. R., Maydew, E. L., & Sparks, H. C. (1999). The role of Big 6 auditors in the credible reporting of accruals. Auditing: A Journal of Practice & Theory, 18(2), 17-34.DOI: https://doi.org/10.2308/aud.1999.18.2.17
30. Francis, J. R., & Wang, D. (2008). The joint effect of investor protection and Big 4 audits on earnings quality around the world. Contemporary Accounting Research, 25(1), 157-191.DOI: https://doi.org/10.1506/car.25.1.6
31. Guénin-Paracini, H., Malsch, B., & Paillé, A. M. (2014). Fear and risk in the audit process. Accounting, Organizations and Society, 39(4), 264-288.DOI: https://doi.org/10.1016/j.aos.2014.02.001
32. Guénin-Paracini, H., Malsch, B., & Tremblay, M. S. (2015). On the operational reality of auditors' independence: Lessons from the field. Auditing: A Journal of Practice & Theory, 34(2), 201-236.DOI: https://doi.org/10.2308/ajpt-50905
33. Hammersley, J. S. (2006). Pattern identification and industry-specialist auditors. The Accounting Review, 81(2), 309-336.DOI: https://doi.org/10.2308/accr.2006.81.2.309
34. Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, 3(4), 305-360. DOI: https://doi.org/10.1016/0304-405X(76)90026-X
35. 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. DOI: https://doi.org/10.2308/jeta-51730
36. KPMG. (2024). KPMG Clara: Smart Audit Platform. KPMG International.Disponible: https://home.kpmg/
37. Lawrence, A., Minutti-Meza, M., & Zhang, P. (2011). Can Big 4 versus non-Big 4 differences in audit-quality proxies be attributed to client characteristics? The Accounting Review, 86(1), 259-286. DOI: https://doi.org/10.2308/accr.00000009
38. López, D. M., & Peters, G. F. (2012). The effect of workload compression on audit quality. Auditing: A Journal of Practice & Theory, 31(4), 139-165. https://doi.org/10.2308/ajpt-10305
39. Moffitt, K. C., Richardson, V. J., Snow, M. S., Weisner, M. M., & Wood, D. A. (2018). Financial Statement Audits: Automation with Algorithms. Brigham Young University.
40. Moore, D. A., Tetlock, P. E., Tanlu, L., & Bazerman, M. H. (2006). Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. Academy of Management Review, 31(1), 10-29.DOI: https://doi.org/10.5465/amr.2006.19379621
41. Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. Journal of Business Ethics, 167, 209-234. DOI: https://doi.org/10.1007/s10551-019-04407-1
42. PCAOB. (2023). 2022 Inspection Observations. Public Company Accounting Oversight Board. Disponible: https://pcaobus.org/
43. Perols, J. L., & Lougee, B. A. (2011). The relation between earnings management and financial statement fraud. Advances in Accounting, 27(1), 39-53.DOI: https://doi.org/10.1016/j.adiac.2010.10.004
44. PwC. (2024). GL.ai: Contract Analysis Platform. PricewaterhouseCoopers.Disponible: https://www.pwc.com/
45. Quick, R., & Warming-Rasmussen, B. (2009). Auditor independence and the provision of non-audit services: Perceptions by German investors. International Journal of Auditing, 13(2), 141-162. DOI: https://doi.org/10.1111/j.1099-1123.2008.00396.x
46. SEC. (2023). Enforcement Actions Against Audit Firms. U.S. Securities and Exchange Commission.Disponible: https://www.sec.gov/
47. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257-285. DOI: https://doi.org/10.1207/s15516709cog1202_4

48. Teoh, S. H., & Wong, T. J. (1993). Perceived auditor quality and the earnings response coefficient. The Accounting Review, 68(2), 346-366.
49. Watkins, A. L., Hillison, W., & Morecroft, S. E. (2004). Audit quality: A synthesis of theory and empirical evidence. Journal of Accounting Literature, 23, 153-193.
50. Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2022). The impact of artificial intelligence and blockchain on the accounting profession. IEEE Access, 8, 110461-110477. DOI: https://doi.org/10.1109/ACCESS.2020.3002316