



Review Paper

Opening the Black Box: A Framework for Auditor Assurance Over the Explainability of AI Decisions in IT Risk Governance¹

Farzad Evani^{*2}  and Donya Kazemi³

Journal of Information System and Technology Auditing
Iranian Information Technology Audit Scientific
Association
Vol. 2, No. 1, Spring & Summer 2026
pp. 159-186

Received: 2026.08.20
Revised: 2026.10.10
Accepted: 2026.10.28

Abstract

As organizations increasingly rely on artificial intelligence (AI) systems for decisions such as fraud detection and credit risk scoring, auditors face a growing challenge: how to provide professional assurance over decisions produced by opaque, “black-box” models. While traditional IT audit frameworks focus mainly on the accuracy and efficiency of system outputs, explainability has received comparatively little attention as an independent dimension of AI risk governance. This conceptual, literature-based study examines the position of AI explainability within IT governance frameworks (COBIT), the ISO/IEC 42001 standard, and international audit assurance standards. The review finds that Iranian domestic literature treats AI mainly as a tool to improve audit accuracy and judgment quality, while international literature and standard-setters are rapidly moving toward requiring explainability and algorithmic accountability. Based on this gap, the study proposes a three-layer conceptual framework — technical transparency, governance documentation, and assurance reporting — to guide auditors’ evaluation of AI explainability, offering a foundation for future empirical research and practical guidance for Iranian IT auditors.

Keywords: Explainable AI (XAI), IT Risk Governance, Audit Assurance, Algorithmic Accountability, AI Governance.

JEL classification: M42, M15, O33.

¹ <https://doi.org/10.22034/JISTA.2026.603372.1107>

¹ Selected paper of the 2nd National Conference on Corporate Governance and Information Technology Risk Management

² Assistant Professor, Department of Accounting, Faculty of Economics and Accounting, Razi University, Kermanshah, Iran. (Corresponding Author). Email: feivani@gmail.com

³ M.A. in Accounting, Faculty of Economics and Accounting, Razi University, Kermanshah, Iran. Email: kazemidonya95@gmail.com

1. INTRODUCTION

Over the past decade, artificial intelligence has moved from a supporting analytical tool to a central driver of organizational decision-making. Machine learning systems are now used for financial fraud detection, credit risk assessment, customer scoring, and audit risk prediction. This expansion, while offering clear gains in speed and accuracy, raises a fundamental question for the audit profession: when an organizational decision emerges from a model whose internal logic even its designers cannot fully explain, on what basis can an auditor provide assurance over the correctness, fairness, and reliability of that decision? This phenomenon — commonly referred to as the “black box” problem — is not merely a technical issue but a governance one. IT governance frameworks such as COBIT, which have traditionally focused on strategic alignment, risk management, and value creation, must now respond to a new dimension: algorithmic explainability. At the regulatory level, ISO/IEC 42001, introduced in late 2023, is the first formal management-system standard dedicated to AI, and comparative research (McIntosh et al., 2024) shows it offers broader coverage of generative-AI opportunities than earlier frameworks such as COBIT and ISO 27001, though gaps remain in risk oversight. International audit standard-setters have also responded: the IAASB’s October 2024 Technology Position committed the board to a gap analysis of existing standards against emerging technologies, followed by a 2025 Technology Quality Management initiative and global roundtables with over 240 stakeholders from six continents. In Iran, by contrast, existing research on AI and auditing has concentrated mainly on improving audit quality, modeling auditor judgment, and analyzing management ability, with little attention to the transparency of the decision-making process itself. This gap motivates the present study, whose central research question is: in the absence of standardized criteria for assessing AI explainability, on what basis can an IT auditor provide valid assurance?



2. MATERIALS AND METHODS

This study adopts a qualitative, conceptual research design based on a systematic review of domestic (Iranian) and international literature. The search strategy combined keywords such as “explainable AI,” “AI governance,” “IT audit,” “COBIT,” and “algorithmic assurance” across Iranian peer-reviewed accounting and IT-management journals and international databases indexed in Scopus and Web of Science, including arXiv, ScienceDirect, and Frontiers. Inclusion criteria required publication in a peer-reviewed journal or conference proceedings and direct relevance to one of three themes: AI explainability, IT governance frameworks, or audit assurance. Selected sources were organized into domestic and international categories to make the gap between the two literatures explicit and were then analyzed and compared to identify the main conceptual and practical divergences. Because the study is conceptual rather than empirical, no primary field data (e.g., survey or interview data) were collected; the objective is to synthesize existing knowledge into a framework capable of guiding future empirical research.

3. RESULTS AND DISCUSSION

The review's first finding confirms a gap between domestic and international literature: while Iranian studies (e.g., Sahat Barmehcheh et al., 2025; Shah Sahebi et al., 2020; Ahmad Hossein et al., 2024) treat AI mainly as a tool for improving audit judgment, management-ability analysis, or bank performance evaluation, international literature and standard-setters are rapidly defining independent frameworks for evaluating explainability and algorithmic accountability. Building on agency theory, the paper argues that AI explainability is a logical extension of the classic audit function: just as independent audit historically reduced information asymmetry between management and owners, evaluating AI explainability can reduce the information asymmetry between opaque algorithmic systems and organizational stakeholders. Based on this theoretical grounding, the study proposes a three-layer conceptual framework for auditing AI explainability, illustrated in Figure 1 and summarized in Table 1.



Layer 1, Technical Transparency, assesses whether technical documentation of the model (model type, training data, and explainability methods such as SHAP or LIME) is available and comprehensible to the auditor; given evidence that even popular post-hoc methods can lack sufficient fidelity in multilevel models (Bhaumik et al., 2022), auditors must also assess the reliability of these explanations, not merely their existence. Layer 2, Governance Documentation, evaluates organizational alignment with governance frameworks such as COBIT and ISO/IEC 42001, including whether model-risk ownership and periodic review procedures are formally defined. Layer 3, Assurance Reporting, addresses how findings are reflected in the auditor's report; in the absence of a dedicated Iranian reporting standard, the study recommends a qualified-disclosure approach similar to that used in auditing complex accounting estimates.

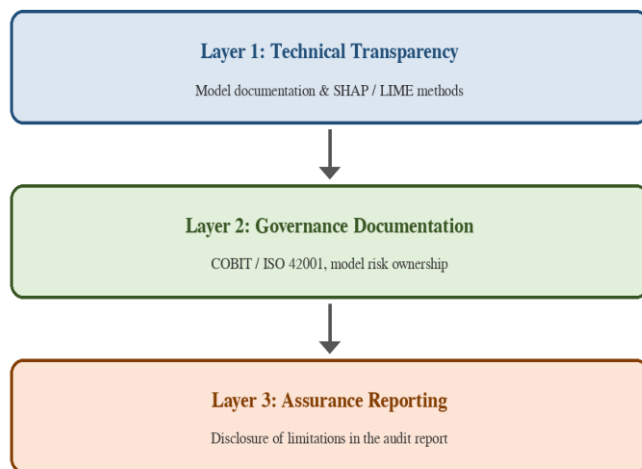


Figure 1. Proposed three-layer framework for auditing AI explainability

A hypothetical illustrative scenario — an Iranian bank applying the framework to an AI-based credit-scoring system — shows how each layer would guide the auditor: verifying whether SHAP-based explanations exist and have been validated (Layer 1); confirming that the risk committee has formally assigned model-risk ownership



and a periodic review cadence (Layer 2); and, where documentation is incomplete, disclosing that limitation explicitly in the audit report rather than remaining silent (Layer 3). This illustrates that, unlike purely technical approaches in the domestic literature, the proposed framework gives auditors a concrete decision path even in the absence of a formal standard.

Table 1

Comparative analysis of governance frameworks by explainability coverage

Framework	Primary Focus	Explainability Coverage	Regulatory Alignment
COBIT 2019	Strategic alignment & IT risk management	Moderate; requires added AI-specific processes	Strongest alignment with EU AI Act
ISO/IEC 42001 (2023)	Dedicated AI management systems	Comprehensive; purpose-built for AI	Broadest coverage of generative-AI opportunities
NIST AI RMF	Technical AI risk management	High technical depth; weaker governance layer	Non-binding, primarily US reference
ISO 27001	Organizational information security	Limited; explainability outside core scope	Gaps in AI-specific risk oversight

Source: compiled by the authors based on McIntosh et al. (2024) and ISACA (2025).

4. CONCLUSION

This study shows that AI explainability, as a dimension independent of model accuracy, remains underexplored in Iran's domestic IT-audit literature, while international literature and professional bodies are rapidly moving toward requiring it. The proposed three-layer framework — technical transparency, governance documentation, and assurance reporting — offers a starting point for future empirical work, including case studies in Iranian organizations and readiness surveys of internal auditors. Practically, audit firms should update internal training on explainable-AI fundamentals, audit committees should treat AI governance as a board-level matter rather than a purely technical



one, and Iranian professional bodies should develop operational guidance for adapting COBIT to AI-governance requirements. As a conceptual review, the study's main limitation is the absence of primary field data; validating the proposed framework empirically is left for future research.

COPYRIGHT



This license allows others to download the works and share them with others as long as they credit them, but they can't change them in any way or use them commercially.



مقاله ترویجی

گشودن جعبه سیاه: چارچوبی برای اطمینان بخشی حساب‌رسان از توضیح پذیری تصمیمات هوش مصنوعی در حاکمیت ریسک فناوری اطلاعات^۱ فرزاد ایوانی^{۲*} و دنیا کاظمی^۳

تاریخ دریافت: ۱۴۰۵/۰۵/۲۹

تاریخ بازنگری: ۱۴۰۵/۰۷/۱۸

تاریخ پذیرش: ۱۴۰۵/۰۸/۰۶

حسابرسی سیستم‌ها و فناوری اطلاعات

انجمن حسابرسی فناوری اطلاعات ایران

سال دوم، پیاپی ۱، بهار و تابستان ۱۴۰۵

صص ۱۵۹ - ۱۸۶

چکیده

این پژوهش با رویکردی مفهومی و مروری، جایگاه توضیح‌پذیری هوش مصنوعی را در حاکمیت ریسک فناوری اطلاعات و اطمینان بخشی حساب‌رسان بررسی می‌کند. با گسترش به کارگیری هوش مصنوعی در فرایندهای سازمانی، از جمله امتیازدهی اعتباری و تشخیص تقلب، حساب‌رسان با این پرسش مواجه‌اند که چگونه می‌توان تصمیماتی را اطمینان بخشی کرد که از دل مدل‌های غیر شفاف (جعبه سیاه) بیرون می‌آیند. داده‌های پژوهش از مرور نظام‌مند ادبیات داخلی و خارجی مرتبط با هوش مصنوعی توضیح‌پذیر، کویت و ایزو ۴۲۰۰۱ گردآوری و با رویکرد تطبیقی بررسی شده است. یافته‌ها نشان می‌دهد ادبیات داخلی عمدتاً بر بهبود دقت حساب‌رسان با کمک هوش مصنوعی متمرکز است و کمتر به توضیح‌پذیری به عنوان بُعدی مستقل از حاکمیت ریسک پرداخته، درحالی‌که نهادهای بین‌المللی به سرعت به سوی الزامی کردن آن حرکت می‌کنند. بر این اساس، چارچوب مفهومی سه‌لایه شامل شفافیت فنی، مستندسازی حاکمیتی و گزارشگری اطمینان بخشی پیشنهاد شده که با سناریویی فرضی در حوزه بانکداری نشان داده شده و می‌تواند مبنای پژوهش‌های تجربی آتی و راهنمای عملیاتی برای حساب‌رسان ایرانی باشد.

واژه‌های کلیدی: هوش مصنوعی توضیح‌پذیر، حاکمیت ریسک فناوری اطلاعات، اطمینان بخشی

حساب‌رسان، پاسخ‌گویی الگوریتمی.

طبقه‌بندی موضوعی: M42, M15, O33.

^۱ <https://doi.org/10.22034/JISTA.2026.603372.1107>

^۲ مقاله منتخب دومین همایش ملی حاکمیت شرکتی و مدیریت ریسک‌های فناوری اطلاعات

^۳ استادیار گروه حسابداری، دانشکده اقتصاد و حسابداری، دانشگاه رازی، کرمانشاه، ایران. (نویسنده مسئول). (feivani@gmail.com)

^۴ کارشناس ارشد حسابداری، دانشکده اقتصاد و حسابداری، دانشگاه رازی، کرمانشاه، ایران. (kazemidonya95@gmail.com)

مقدمه

در یک دهه اخیر، هوش مصنوعی از یک ابزار کمکی در تحلیل داده‌ها، به یکی از فناوری‌های مهم در فرایندهای تصمیم‌گیری و مدیریت سازمانی تبدیل شده است. استفاده از روش‌های یادگیری ماشین و سایر فناوری‌های هوش مصنوعی در فرایندهای حسابرسی و ارزیابی مدل‌های تصمیم‌گیری، ظرفیت‌هایی برای افزایش کارایی و کیفیت تحلیل‌ها ایجاد کرده است؛ هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی نیز بر ظرفیت فناوری برای ارتقای کیفیت کار حسابرسی و اطمینان بخشی تأکید کرده است (هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی^۱، ۲۰۲۴: ۱).

با گسترش استفاده از مدل‌های پیچیده هوش مصنوعی، مسئله شفافیت و قابلیت توضیح تصمیم‌های الگوریتمی نیز اهمیت بیشتری یافته است. برخی مدل‌های یادگیری ماشین از نظر نحوه تبدیل ورودی‌ها به خروجی‌ها، برای کاربران و حتی ارزیابان فنی به سادگی قابل تفسیر نیستند و از این منظر، موضوع «توضیح‌پذیری» و «قابلیت حسابرسی» مدل‌های هوش مصنوعی اهمیت ویژه‌ای پیدا می‌کند. پژوهش بهاومیک، دی و کایال^۲ (۲۰۲۲: ۱۲) با تمرکز بر حسابرسی مدل‌های چندسطحی، ارزیابی مدل، تبعیض، شفافیت و توضیح‌پذیری را از ابعاد مهم حسابرسی مدل‌های هوش مصنوعی معرفی می‌کند و استفاده از روش‌هایی مانند شپ^۳ و لایم^۴ را برای ارزیابی توضیح‌پذیری مورد بررسی قرار می‌دهد.

در حوزه حاکمیت فناوری اطلاعات نیز چارچوب‌هایی مانند کویت^۵ باهدف ایجاد سازوکارهای مناسب برای حاکمیت و مدیریت فناوری اطلاعات مورد استفاده قرار گرفته‌اند. با گسترش کاربرد هوش مصنوعی، این چارچوب‌ها با ضرورت توجه بیشتر به فرصت‌ها، ریسک‌ها و الزامات حاکمیت مرتبط با سامانه‌های هوش مصنوعی مواجه شده‌اند (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی^۶، ۲۰۲۵: ۳؛ مک‌اینش و همکاران^۷، ۲۰۲۴: ۲-۱).

¹ International Auditing and Assurance Standards Board (IAASB)

² Bhaumik, Dey & Kayal

³ SHapley Additive exPlanations (SHAP)

⁴ Local Interpretable Model-agnostic Explanations (LIME)

⁵ Control Objectives for Information and Related Technologies (COBIT)

⁶ Information Systems Audit and Control Association (ISACA)

⁷ McIntosh et al.



چهار چارچوب شامل چارچوب امنیت سایبری ان‌آی‌اس‌تی^۱، کویت ۲۰۱۹، ایزو ۲۷۰۰۱ و ایزو ۴۲۰۰۱ را از منظر آمادگی برای استفاده از مدل‌های زبانی بزرگ بررسی کرده است. نتایج این پژوهش نشان می‌دهد که ایزو ۴۲۰۰۱ در میان چارچوب‌های بررسی شده، جامع‌ترین ظرفیت را برای تسهیل فرصت‌های مرتبط با مدل‌های زبانی بزرگ داشته است؛ باین‌حال، همچنان شکاف‌هایی در نظارت بر ریسک‌های مرتبط با این فناوری وجود دارد.

در سطح استاندارد گذاری بین‌المللی، استاندارد ایزو ۴۲۰۰۱ در دسامبر ۲۰۲۳ منتشر شد. این استاندارد، الزامات مربوط به ایجاد، استقرار، نگهداری و بهبود مستمر «سیستم مدیریت هوش مصنوعی»^۲ را مشخص می‌کند و توسط سازمان بین‌المللی استاندارد سازی^۳ به‌عنوان نخستین استاندارد سیستم مدیریت هوش مصنوعی معرفی شده است. این استاندارد همچنین بر مدیریت ریسک‌ها و فرصت‌های مرتبط با هوش مصنوعی، شفافیت و استفاده مسئولانه از این فناوری تأکید دارد (سازمان بین‌المللی استاندارد سازی، ۲۰۲۳؛ ۱؛ مک‌اینتاش و همکاران، ۲۰۲۴؛ ۱).

در ایران نیز پژوهش‌های انجام‌شده در زمینه کاربرد هوش مصنوعی در حسابداری و حسابرسی (شاه‌صاحبی و همکاران، ۱۳۹۹؛ سلیمانی‌امیری و محمدی‌سه‌دران، ۱۴۰۳؛ صحت‌برمجه و همکاران، ۱۴۰۵) بیشتر به بررسی نقش این فناوری در بهبود فرایندهای حسابرسی، افزایش کیفیت قضاوت حساب‌رسان و تحلیل برخی ابعاد مرتبط با توانایی مدیریت پرداخته‌اند. با وجود این، در مطالعات بررسی‌شده، موضوع شفافیت و توضیح‌پذیری نحوه تصمیم‌گیری مدل‌های هوش مصنوعی و ارتباط آن با فرایندهای اطمینان‌بخشی در حسابرسی فناوری اطلاعات کمتر به طور مستقیم مورد توجه قرار گرفته است. به نظر می‌رسد بخش قابل توجهی از ادبیات داخلی همچنان بر استفاده از هوش مصنوعی برای انجام یا تقویت فعالیت‌های حسابرسی متمرکز است، در حالی که در ادبیات بین‌المللی، توجه بیشتری به بررسی و ارزیابی خود سامانه‌های هوش مصنوعی، از جمله سازوکارهای حاکمیت، مدیریت ریسک و توضیح‌پذیری آن‌ها، شکل گرفته است. از این منظر، بررسی هم‌زمان توضیح‌پذیری مدل‌های هوش مصنوعی و ارتباط آن با حاکمیت و مدیریت ریسک فناوری اطلاعات و فرایند

¹ NIST Cybersecurity Framework (NIST CSF)

² Artificial Intelligence Management System

³ International Organization for Standardization (ISO)



اطمینان بخشی حسابرسی، از جمله موضوعاتی است که در پژوهش های داخلی کمتر مورد توجه قرار گرفته و می تواند زمینه مناسبی برای مطالعات بیشتر فراهم کند.

از این منظر، مسئله اساسی پژوهش حاضر آن است که با افزایش استفاده از سامانه های هوش مصنوعی در سازمان ها، حسابرس فناوری اطلاعات چگونه می تواند درباره شفافیت، قابلیت توضیح و کنترل ریسک های تصمیم های الگوریتمی اطمینان بخشی ارائه کند؛ چرا که پیچیدگی این مدل ها، ارزیابی منطق تصمیم گیری و تشخیص سوگیری را دشوار می سازد.

در سطح حرفه حسابرسی نیز تحولات فناوری مورد توجه نهادهای استاندارد گذار قرار گرفته است. هیئت استانداردهای بین المللی حسابرسی و اطمینان بخشی در بیانیه موضع فناوری خود، هم گرایی حسابرسی، اطمینان بخشی و فناوری را مبنای رویکرد خود دانسته و هشت اقدام راهبردی برای تسهیل و تشویق به کارگیری فناوری از سوی حسابسان تعیین کرده است؛ از جمله رفع موانع واقعی یا ادراک شده در استانداردها برای استفاده از فناوری و تدوین الزامات و مواد کاربردی مرتبط با استفاده از آن در اجرای کار (هیئت استانداردهای بین المللی حسابرسی و اطمینان بخشی، ۲۰۲۴: ۱).

در حوزه حسابرسی داخلی نیز «استانداردهای جهانی حسابرسی داخلی»^۱ در ژانویه ۲۰۲۴ منتشر شدند. این استانداردها، افزون بر دانش مربوط به فناوری اطلاعات، آشنایی با ابزارها و فنون گردآوری، تحلیل و ارزیابی داده را جزو شایستگی های لازم حسابسان داخلی می شمارند (انجمن حسابسان داخلی^۲، ۲۰۲۴: ۲، ۲۶).

براین اساس، سرعت گسترش و استفاده از فناوری های هوش مصنوعی در سازمان ها، در کنار نیاز به ایجاد سازوکارهای مناسب برای حاکمیت، مدیریت ریسک، شفافیت و توضیح پذیری این فناوری ها، یکی از مهم ترین زمینه های شکل گیری پژوهش حاضر است. در این راستا، پژوهش حاضر با ارائه یک چارچوب مفهومی سه لایه تلاش می کند توضیح پذیری هوش مصنوعی را صرفاً یک ویژگی فنی در نظر نگیرد، بلکه آن را در ارتباط با حاکمیت و مدیریت ریسک فناوری اطلاعات و همچنین فرایند اطمینان بخشی حسابرسی فناوری اطلاعات بررسی

¹ Global Internal Audit Standards

² Institute of Internal Auditors (IIA)



کند. این رویکرد با در نظر گرفتن شرایط و الزامات سازمانی ایران، می‌تواند تصویری منسجم‌تر از جایگاه توضیح‌پذیری در فرایند ارزیابی و اطمینان‌بخشی سامانه‌های هوش مصنوعی ارائه دهد.

مبانی نظری

هوش مصنوعی توضیح‌پذیر

هوش مصنوعی توضیح‌پذیر^۱، به مجموعه‌ای از روش‌ها و رویکردهایی اشاره دارد که هدف آن‌ها افزایش قابلیت درک و تفسیر فرایند تصمیم‌گیری مدل‌های هوش مصنوعی برای کاربران انسانی است. در ادبیات مربوط به توضیح‌پذیری، میان روش‌های مبتنی بر مدل‌های ذاتاً تفسیرپذیر و روش‌های توضیح پسینی، تمایز قائل می‌شود. در روش‌های پسینی، رفتار مدل‌های پیچیده پس از آموزش با استفاده از ابزارهایی مانند شپ و لایم مورد تحلیل قرار می‌گیرد. علاوه بر این، روش‌های تبیین ضدواقعی^۲ نیز با تمرکز بر تغییرات لازم در ویژگی‌های ورودی برای تغییر خروجی مدل، نوع دیگری از توضیح برای تصمیم‌های الگوریتمی فراهم می‌کنند. این رویکردها از منظر حسابرسی اهمیت دارند؛ زیرا حسابرس نه تنها به خروجی مدل، بلکه به شواهدی درباره نحوه شکل‌گیری آن خروجی نیاز دارد.

بهاومیک و همکاران (۲۰۲۲: ۱۲) با ارائه چارچوبی برای حسابرسی مدل‌های چندسطحی، سه حوزه مدل، تبعیض و شفافیت و توضیح‌پذیری را برای ارزیابی حسابرسی مدل‌ها مورد توجه قرار داده‌اند. آنان روش‌هایی مانند شپ و لایم را در ارزیابی توضیح‌پذیری بررسی کرده و نشان داده‌اند که این روش‌ها در مدل‌های مورد مطالعه، در برخی جنبه‌های تفسیر اهمیت ویژگی‌ها و سهم ویژگی‌ها با محدودیت‌هایی مواجه هستند (بهاومیک و همکاران، ۲۰۲۲: ۱۹-۲۰). این یافته نشان می‌دهد که استفاده از ابزارهای توضیح‌پذیری به تنهایی برای اطمینان از قابلیت تفسیر یک مدل کافی نیست و خود روش‌های توضیح‌پذیری نیز می‌توانند موضوع ارزیابی و حسابرسی قرار گیرند.

¹ Explainable Artificial Intelligence (XAI)

² Counterfactual Explanations



چارچوب‌های حاکمیت فناوری اطلاعات و هوش مصنوعی

حاکمیت فناوری اطلاعات، از سازوکارهای مهم برای تعیین مسئولیت‌ها، مدیریت ریسک و همسویی فناوری با اهداف سازمانی است. چارچوب کوییت که توسط انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی توسعه یافته، از شناخته شده ترین چارچوب‌های حاکمیت فناوری اطلاعات است و با گسترش هوش مصنوعی، می‌تواند برای پاسخگویی، شفافیت، مدیریت ریسک و نظارت بر این سامانه‌ها نیز به کار رود (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی، ۲۰۲۵: ۳).

انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی (۲۰۲۵: ۳، ۱۳-۱۴) در راهنمای «بهره‌گیری از کوییت برای حاکمیت مؤثر سامانه‌های هوش مصنوعی» توضیح می‌دهد که کوییت می‌تواند برای حاکمیت این سامانه‌ها در طول چرخه عمرشان، شامل پاسخگویی، آزمون، اعتبارسنجی و پایش مستمر، به کار رود. در کنار کوییت، استاندارد ایزو ۴۲۰۰۱ به طور اختصاصی برای سیستم مدیریت هوش مصنوعی طراحی شده و چارچوبی برای ایجاد، استقرار، نگهداری و بهبود این سیستم فراهم می‌کند (سازمان بین‌المللی استانداردسازی، ۲۰۲۳). استفاده هم‌زمان از اصول حاکمیتی کوییت و الزامات ایزو ۴۲۰۰۱ می‌تواند مبنای مناسبی برای ارزیابی حاکمیتی سامانه‌های هوش مصنوعی فراهم کند.

در سطح تنظیم‌گری نیز الزامات مرتبط با شفافیت، پاسخگویی، مدیریت ریسک و نظارت بر سامانه‌های هوش مصنوعی اهمیت فزاینده‌ای یافته‌اند. چونگ^۱ (۲۰۲۴: ۱-۲) در یک مرور درباره شفافیت و پاسخگویی در سامانه‌های هوش مصنوعی، چالش‌های موجود را در چهار حوزه اصلی شامل رویکردهای فنی، چارچوب‌های حقوقی و تنظیم‌گری، ملاحظات اخلاقی و اجتماعی و رویکردهای میان‌رشته‌ای و چندذی‌نفعی طبقه‌بندی می‌کند. این طبقه‌بندی نشان می‌دهد که ارزیابی سامانه‌های هوش مصنوعی صرفاً یک مسئله فنی نیست و ابعاد حقوقی، اخلاقی، اجتماعی و حاکمیتی نیز باید در ارزیابی آن‌ها مورد توجه قرار گیرد.

¹ Cheong



نظریه اطمینان بخشی حسابرسی در بستر الگوریتمی

اطمینان بخشی حسابرسی بر فراهم کردن شواهد کافی برای اظهار نظر حرفه‌ای استوار است. با ورود هوش مصنوعی، دامنه اطلاعات لازم فراتر از اسناد مالی سنتی می‌رود و شامل داده‌های مورد استفاده، طراحی و عملکرد مدل، فرایند اعتبارسنجی و شواهد مربوط به نحوه تولید خروجی می‌شود.

از این منظر، توضیح پذیری می‌تواند بُعدی مهم در ارزیابی سامانه‌های هوش مصنوعی باشد؛ زیرا حسابرس باید تا حد امکان درباره منطق تصمیم‌گیری و ریسک‌های خروجی مدل قضاوت کند (هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی، ۲۰۲۴: ۱؛ مک‌اینش و همکاران، ۲۰۲۴: ۲-۱).

بهاومیک و همکاران (۲۰۲۲: ۱۲، ۲۰) نیز نشان می‌دهند ارزیابی مدل باید فراتر از سنجش صرف عملکرد پیش‌بینی آن باشد. همچنین چونگ (۲۰۲۴: ۱-۲) نشان می‌دهد که شفافیت و پاسخگویی در سامانه‌های هوش مصنوعی با مجموعه‌ای از چالش‌های فنی، حقوقی، اخلاقی و اجتماعی همراه است؛ بنابراین، در بستر حسابرسی فناوری اطلاعات، ارزیابی توضیح‌پذیری را می‌توان بخشی از یک رویکرد گسترده‌تر برای بررسی شفافیت، پاسخگویی و مدیریت ریسک سامانه‌های هوش مصنوعی در نظر گرفت.

واکنش نهادهای بین‌المللی تنظیم‌گری حسابرسی

هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی در بیانیه موضع فناوری خود، افزون بر این، بر ایجاد توازن در اشاره به فرصت‌ها و ریسک‌های فناوری، هم‌راستایی با اصول اخلاقی و فرایند مستمر پایش روندهای فناورانه تأکید کرده است (هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی، ۲۰۲۴: ۱).

این روند در سال‌های ۲۰۲۵ و ۲۰۲۶ نیز با برگزاری میزگردهای جهانی «مدیریت کیفیت فناوری» ادامه یافته است. محور این میزگردها، تأثیر فناوری‌های نوظهور، از جمله هوش مصنوعی، بر حسابرسی و اطمینان بخشی و ضرورت حفظ اعتماد در استفاده از ابزارهای فناورانه بوده است (هیئت استانداردهای بین‌المللی حسابرسی و اطمینان بخشی، ۲۰۲۶).



این تحولات نشان می‌دهد که استفاده از هوش مصنوعی در حسابرسی صرفاً یک مسئله فنی مربوط به عملکرد ابزارها نیست، بلکه به موضوعاتی همچون مدیریت کیفیت، حاکمیت، مدیریت ریسک و پاسخگویی نیز مرتبط شده است. از این منظر، توضیح‌پذیری و شفافیت مدل‌های هوش مصنوعی می‌توانند به عنوان بخشی از مسائل حاکمیتی و کیفیتی مورد توجه حسابرسی فناوری اطلاعات قرار گیرند.

نظریه نمایندگی و ضرورت حاکمیت هوش مصنوعی

بر اساس نظریه نمایندگی، جدایی مالکیت از مدیریت می‌تواند به شکل‌گیری تضاد منافع و عدم تقارن اطلاعاتی میان مالکان و مدیران منجر شود و سازوکارهایی مانند نظارت و حسابرسی برای کاهش این مسائل اهمیت پیدا می‌کنند (جنسن و مک‌لینگ^۱، ۱۹۷۶: ۳۰۶، ۳۰۸).

ورود هوش مصنوعی به فرایندهای تصمیم‌گیری سازمانی می‌تواند شکل جدیدی از عدم تقارن اطلاعاتی ایجاد کند؛ زیرا ذی‌نفعان ممکن است درباره نحوه شکل‌گیری تصمیم‌های الگوریتمی اطلاعات کافی نداشته باشند. در چنین شرایطی، شفافیت و توضیح‌پذیری می‌تواند این عدم تقارن را کاهش داده و امکان نظارت را افزایش دهند (چونگ، ۲۰۲۴: ۲).

از این منظر، ارزیابی توضیح‌پذیری هوش مصنوعی را می‌توان امتدادی از منطق نظارتی نظریه نمایندگی در محیط الگوریتمی دانست: همان‌گونه که حسابرسی مستقل با کاهش عدم تقارن اطلاعاتی میان مدیریت و ذی‌نفعان یک سازوکار نظارتی است، ارزیابی شفافیت و توضیح‌پذیری سامانه‌های هوش مصنوعی نیز می‌تواند نظارت مؤثرتری بر تصمیم‌های الگوریتمی فراهم کند.

در ایران، بخشی از پژوهش‌های حوزه هوش مصنوعی و حسابداری و حسابرسی بر کاربرد این فناوری در فرایندهای حسابرسی، قضاوت حسابرسی، تحلیل توانایی مدیریت و ارزیابی عملکرد تمرکز داشته‌اند؛ از جمله سلیمانی امیری و محمدی سهدران (۱۴۰۳: ۱) درباره حسابرسی در عصر هوش مصنوعی، صحت برمجه، شماخی و یلفانی (۱۴۰۵: ۱۹۱) درباره الگوی قضاوت حسابرسی مبتنی بر هوش مصنوعی، شاه صاحبی، دارابی و حمیدیان (۱۳۹۹: ۷۵) درباره توانایی مدیریت، و احمد حسین، زواری رضایی و آشتاب (۱۴۰۳: ۷۴) درباره عملکرد مالی بانک‌ها.

¹ Jensen & Meckling



با این حال، تمرکز این مطالعات عمدتاً بر کاربرد هوش مصنوعی به عنوان ابزار تحلیل، پیش‌بینی یا تصمیم‌گیری بوده و توضیح‌پذیری و شفافیت فرایند تصمیم‌گیری مدل‌ها به عنوان موضوعی مستقل مورد توجه قرار نگرفته است. پژوهش حاضر با انتقال تمرکز از «استفاده از هوش مصنوعی در حسابرسی» به «ارزیابی قابلیت توضیح و شفافیت آن از منظر حسابرسی فناوری اطلاعات»، این بعد کمتر مورد توجه را بررسی می‌کند.

در حوزه حاکمیت فناوری اطلاعات نیز رونقی و محمودی (۱۳۹۴: ۶۱۵) رابطه حاکمیت فناوری اطلاعات و حاکمیت شرکتی را در سازمان‌های دولتی بررسی کرده و رونقی، فرهادی و خردمند (۱۳۹۷: ۹) به ارزیابی حاکمیت فناوری اطلاعات با رویکرد سلسله‌مراتبی خاکستری پرداخته‌اند. این مطالعات اهمیت حاکمیت فناوری اطلاعات را نشان می‌دهند، اما به شفافیت و توضیح‌پذیری مدل‌های هوش مصنوعی نپرداخته‌اند.

در ادبیات بین‌المللی نیز بهاومیک و همکاران (۲۰۲۲: ۱۲) رابطه میان حسابرسی مدل‌ها و توضیح‌پذیری را بررسی کرده و مک‌این‌تاش و همکاران (۲۰۲۴: ۱) ظرفیت چارچوب‌های حاکمیتی، از جمله کویت و ایزو ۴۲۰۰۱، را در مواجهه با ریسک‌های مدل‌های زبانی بزرگ تحلیل کرده‌اند؛ ترکیب این دو جریان پژوهشی مبنایی برای چارچوب مفهومی پژوهش حاضر فراهم می‌کند.

باتوجه به ماهیت کیفی و مفهومی پژوهش حاضر، اهداف پژوهش به جای تدوین فرضیه‌های علی، به صورت زیر تعیین شده‌اند:

- ۱) شناسایی جایگاه توضیح‌پذیری هوش مصنوعی در ادبیات داخلی و خارجی حسابرسی فناوری اطلاعات؛
- ۲) تحلیل تطبیقی ظرفیت چارچوب‌های حاکمیتی موجود، از جمله کویت، ایزو ۴۲۰۰۱ و چارچوب مدیریت ریسک هوش مصنوعی ان‌آی‌اس تی^۱، برای پوشش ابعاد مرتبط با توضیح‌پذیری، شفافیت، پاسخگویی و مدیریت ریسک هوش مصنوعی؛
- ۳) ارائه یک چارچوب مفهومی برای هدایت ارزیابی حسابرسی فناوری اطلاعات از توضیح‌پذیری سامانه‌های هوش مصنوعی در سازمان‌های ایرانی.

¹ National Institute of Standards and Technology AI Risk Management Framework (NIST AI RMF)



روش‌شناسی پژوهش

این پژوهش با رویکردی کیفی و مفهومی و با استفاده از مرور و تحلیل تطبیقی ادبیات داخلی و بین‌المللی انجام شده است. راهبرد جست‌وجوی منابع بر اساس ترکیبی از کلیدواژه‌های مرتبط با «هوش مصنوعی توضیح‌پذیر»، «حاکمیت هوش مصنوعی»^۱، «حسابرسی فناوری اطلاعات»^۲، «کویت»^۳، «ایزو ۴۲۰۰۱»^۴، «چارچوب مدیریت ریسک هوش مصنوعی نیست» و «اطمینان بخشی الگوریتمی» در منابع علمی و حرفه‌ای مرتبط شکل گرفت.

منابع داخلی از میان مطالعات منتشرشده در حوزه حسابداری، حسابرسی، مدیریت فناوری اطلاعات و حاکمیت فناوری اطلاعات انتخاب شدند. برای بررسی ادبیات بین‌المللی نیز از پایگاه‌ها و ناشران علمی مرتبط از جمله اسکوپوس^۳، وب آو ساینس^۴، ساینس دایرکت^۵، فرانتیرز^۶ و آرکایو^۷ استفاده شد. جست‌وجو در بازه ژانویه ۲۰۱۸ تا ژوئن ۲۰۲۵ انجام شد. کلیدواژه‌های اصلی شامل ترکیب‌های مربوط به ترکیب‌های مربوط به هوش مصنوعی توضیح‌پذیر^۸، حاکمیت فناوری اطلاعات^۹، اطمینان بخشی حسابرسی^{۱۰} و مدل‌های جعبه سیاه^{۱۱} با عملگرهای بولی بود.

معیار اصلی ورود منابع، ارتباط مستقیم آن‌ها با دست‌کم یکی از سه محور اصلی پژوهش شامل توضیح‌پذیری هوش مصنوعی، حاکمیت فناوری اطلاعات و هوش مصنوعی، یا اطمینان بخشی حسابرسی در محیط‌های الگوریتمی بود. معیارهای خروج نیز شامل حذف مطالعات تکراری، منابع فاقد کیفیت علمی کافی، آثار نامرتبط با محورهای پژوهش و مطالعاتی بود که صرفاً به کاربرد ابزاری هوش مصنوعی در حسابرسی پرداخته و به جنبه‌های توضیح‌پذیری و حاکمیت توجهی نداشتند. در مجموع حدود ۱۵۰ منبع بازبایی و پس از غربالگری، ۱۸ منبع برای تحلیل نهایی انتخاب شد. پس از بررسی اولیه، مطالعات مرتبط از منظر موضوع، هدف، رویکرد و یافته‌های اصلی مقایسه شدند. در بخش بین‌المللی، علاوه بر مطالعات دانشگاهی، اسناد

¹ AI Governance

² IT Auditing

³ Scopus

⁴ Web of Science

⁵ ScienceDirect

⁶ Frontiers

⁷ arXiv

⁸ explainable AI / XAI

⁹ IT governance

¹⁰ Audit assurance

¹¹ Black-box models



حرفه‌ای مرتبط با حاکمیت هوش مصنوعی، از جمله کوبیت، ایزو ۴۲۰۰۱ و اسناد هیئت استانداردهای بین‌المللی حسابرسی، نیز موردتوجه قرار گرفتند تا ادبیات دانشگاهی با جهت‌گیری‌های حرفه‌ای مقایسه شود (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی، ۲۰۲۵: ۳؛ مک‌اینتاش و همکاران، ۲۰۲۴: ۴).

جامعه مورد بررسی، ادبیات علمی و حرفه‌ای مرتبط با موضوع است و ابزار گردآوری داده‌ها، مرور و مقایسه تطبیقی محتوای منابع منتخب است. پژوهش حاضر ماهیتی مفهومی-مروری دارد؛ ازاین‌رو هدف آن آزمون تجربی روابط علی نیست، بلکه شناسایی شکاف پژوهشی و ارائه چارچوبی برای هدایت مطالعات آینده است.

یافته‌های پژوهش

یافته‌های حاصل از بررسی و مقایسه تطبیقی ادبیات نشان می‌دهد میان ادبیات داخلی مرتبط با کاربرد هوش مصنوعی در حسابداری و حسابرسی، و ادبیات بین‌المللی مرتبط با توضیح‌پذیری و حاکمیت مدل‌ها، تفاوت‌های آشکاری وجود دارد: در مطالعات داخلی، هوش مصنوعی عمدتاً به‌عنوان ابزار تحلیل و بهبود فرایندها موردتوجه قرار گرفته است؛ در مقابل، بخشی از ادبیات بین‌المللی مستقیماً به حسابرسی مدل، توضیح‌پذیری و حاکمیت این سامانه‌ها پرداخته است (بهاومیک و همکاران، ۲۰۲۲: ۱۲؛ مک‌اینتاش و همکاران، ۲۰۲۴: ۱).

بر اساس این یافته و با تلفیق مباحث مربوط به توضیح‌پذیری، حاکمیت فناوری و اطمینان‌بخشی حسابرسی، پژوهش حاضر، یک چارچوب مفهومی سه‌لایه برای ارزیابی حسابرسی از توضیح‌پذیری هوش مصنوعی پیشنهاد می‌کند.

لایه نخست: شفافیت فنی

در این لایه، حسابرس بررسی می‌کند که آیا اطلاعات و مستندات لازم درباره مدل، داده‌های مورداستفاده، نحوه آموزش و اعتبارسنجی مدل و روش‌های توضیح‌پذیری مورداستفاده در دسترس است یا خیر. در مدل‌های پیچیده، روش‌هایی مانند شپ و لایم می‌توانند برای تولید توضیحات پسینی مورداستفاده قرار گیرند؛ بااین‌حال، استفاده از چنین روش‌هایی نباید به معنای پذیرش خودکار توضیحات تولیدشده به‌عنوان شواهد کافی تلقی شود.



بهاومیک و همکاران (۲۰۲۲: ۲۰) نشان می دهند که روش های توضیح پذیری می تواند خود با محدودیت هایی مواجه باشند و بنابراین ارزیابی کیفیت و قابلیت اتکای توضیحات باید بخشی از فرایند ارزیابی مدل باشد. بر این اساس، حسابرس باید علاوه بر بررسی وجود مستندات توضیح پذیری، قابلیت اتکای روش مورد استفاده و محدودیت های آن را نیز مورد توجه قرار دهد.

لایه دوم: مستندسازی حاکمیتی

در لایه دوم، تمرکز از ویژگی های فنی مدل به سازوکارهای حاکمیتی و سازمانی منتقل می شود. حسابرس بررسی می کند که آیا سازمان فرایندهای مشخصی برای طراحی، توسعه، آموزش، استقرار، اعتبارسنجی، پایش و بازبینی مدل های هوش مصنوعی دارد و آیا مسئولیت ها و اختیارات مربوط به مدیریت ریسک مدل به صورت مشخص تعیین شده اند.

چارچوب کوبیت می تواند از منظر حاکمیت و مدیریت فناوری اطلاعات در این ارزیابی مورد استفاده قرار گیرد. انجمن حسابرسی و کنترل سیستم های اطلاعاتی (۲۰۲۵: ۳، ۱۳) نیز ظرفیت کوبیت را برای پشتیبانی از حاکمیت و مدیریت سامانه های هوش مصنوعی و موضوعاتی مانند پاسخگویی، مدیریت ریسک و کنترل های مرتبط با چرخه عمر فناوری مورد توجه قرار داده است.

در کنار کوبیت، ایزو ۴۲۰۰۱ به طور اختصاصی برای سیستم مدیریت هوش مصنوعی طراحی شده و چارچوبی برای ایجاد، استقرار، نگهداری و بهبود سیستم مدیریت هوش مصنوعی فراهم می کند (سازمان بین المللی استانداردسازی، ۲۰۲۳). بنابراین، استفاده هم زمان از اصول حاکمیت کوبیت و الزامات سیستم مدیریت هوش مصنوعی ایزو ۴۲۰۰۱ می تواند مبنای مناسبی برای ارزیابی حاکمیتی سامانه های هوش مصنوعی فراهم کند.

لایه سوم: گزارشگری اطمینان بخشی

در لایه سوم، نتایج ارزیابی فنی و حاکمیتی به فرایند گزارشگری حسابرسی منتقل می شود. حسابرس باید بتواند محدودیت های شناسایی شده در زمینه مستندات، شفافیت، قابلیت توضیح یا کنترل های حاکمیتی مدل را در چارچوب مناسب گزارشگری منعکس کند.



جدول ۱

چارچوب سه‌لایه پیشنهادی برای ارزیابی حسابرسی از توضیح‌پذیری هوش مصنوعی

لایه	مؤلفه‌های اصلی	پرسش کلیدی حسابرس
اول: شفافیت فنی	نوع مدل، داده‌های مورد استفاده، آموزش و اعتبارسنجی، شپ/لایم و محدودیت‌های توضیح	آیا مدل و روش توضیح‌پذیری مستند، قابل فهم و از نظر فنی قابل اتکا هستند؟
دوم: مستندسازی حاکمیتی	کوبیت، ایزو ۴۲۰۰۱، نقش‌ها و مسئولیت‌ها، مالکیت ریسک مدل، پایش مستمر	آیا مسئولیت‌ها و کنترل‌های حاکمیتی مدل به صورت رسمی تعریف و پایش می‌شوند؟
سوم: گزارشگری اطمینان‌بخشی	یافته‌های حسابرسی، محدودیت‌های توضیح‌پذیری، گزارش نتایج و توصیه‌های اصلاحی	آیا محدودیت‌های مهم در شفافیت و توضیح‌پذیری در گزارش مربوط منعکس شده‌اند؟

در این پژوهش، افشای محدودیت‌های شناسایی شده به عنوان یک پیشنهاد مفهومی مطرح می‌شود، نه به عنوان یک الزام استاندارد موجود؛ بنابراین، پیشنهاد نمی‌شود که حسابرس الزاماً از قالب گزارشگری حسابرسی صورت‌های مالی برای گزارش توضیح‌پذیری هوش مصنوعی استفاده کند؛ بلکه پیشنهاد می‌شود در خدمات اطمینان‌بخشی یا حسابرسی فناوری اطلاعات، محدودیت‌های مهم مربوط به قابلیت توضیح و اتکاپذیری مدل به صورت شفاف در گزارش مربوط منعکس شود. بر اساس یافته‌های حاصل از مقایسه تطبیقی ادبیات، چارچوب سه‌لایه پیشنهادی برای ارزیابی حسابرسی توضیح‌پذیری هوش مصنوعی در جدول ۱ ارائه شده است.



جدول ۲

تحلیل تطبیقی چارچوب های حاکمیتی

چارچوب	کانون اصلی	پوشش مرتبط باهوش مصنوعی	کاربرد پیشنهادی در حسابرسی
کویت ۲۰۱۹	حاکمیت و مدیریت فناوری اطلاعات	حاکمیت، مسئولیت پذیری، کنترل و مدیریت ریسک	ارزیابی ساختار حاکمیتی و کنترل های سازمانی
ایزو ۴۲۰۰۱	سیستم مدیریت هوش مصنوعی	تمرکز اختصاصی بر مدیریت و حاکمیت هوش مصنوعی	ارزیابی سیستم مدیریت هوش مصنوعی و چرخه بهبود
چارچوب مدیریت ریسک هوش مصنوعی نیست	مدیریت ریسک هوش مصنوعی	شناسایی، اندازه گیری و مدیریت ریسک هوش مصنوعی	ساختاردهی ارزیابی ریسک مدل
ایزو ۲۷۰۰۱	امنیت اطلاعات	پوشش عمومی امنیت و کنترل اطلاعات	کنترل های امنیتی مرتبط با داده و سامانه

منبع: تدوین شده توسط نویسندگان بر اساس مک اینتاش و همکاران (۲۰۲۴)، انجمن حسابرسی و کنترل سیستم های اطلاعاتی (۲۰۲۵) و سازمان بین المللی استانداردسازی (۲۰۲۳).

جدول ۳

سیاه اولیه ارزیابی توضیح پذیری

لایه	شاهد مورد انتظار	پرسش ارزیابی
شفافیت فنی	گزارش فنی مدل	آیا نوع مدل، داده های مورد استفاده و روش توضیح پذیری مستند شده است؟
شفافیت فنی	گزارش اعتبارسنجی	آیا عملکرد و محدودیت های روش توضیح پذیری ارزیابی و مستند شده است؟
مستندسازی حاکمیتی	سند نقش ها و مسئولیت ها	آیا مالکیت و مدیریت ریسک مدل به طور رسمی مشخص شده است؟
مستندسازی حاکمیتی	رویه نظارت مستمر	آیا فرایند مشخصی برای بازبینی دوره ای عملکرد و ریسک مدل وجود دارد؟
گزارشگری اطمینان بخشی	گزارش نهایی ارزیابی	آیا محدودیت های مهم شناسایی شده در توضیح پذیری در گزارش مربوط منعکس شده اند؟

منبع: تدوین شده توسط نویسندگان



سناریوی کاربردی فرضی: به کارگیری چارچوب در یک بانک ایرانی

برای نشان دادن نحوه عملیاتی شدن چارچوب سه لایه، یک سناریوی فرضی و غیرواقعی در نظر گرفته می‌شود. فرض کنید یک بانک ایرانی از یک سامانه مبتنی بر یادگیری ماشین برای امتیازدهی اعتباری متقاضیان تسهیلات خرد استفاده می‌کند.

در لایه نخست، یعنی شفافیت فنی، حسابرس فناوری اطلاعات مستندات فنی سامانه را بررسی می‌کند و مشخص می‌سازد که نوع مدل، داده‌های آموزشی، فرایند اعتبارسنجی و روش توضیح‌پذیری مورد استفاده مستند شده‌اند یا خیر. در صورتی که از روش‌هایی مانند شپ استفاده شده باشد، حسابرس باید علاوه بر بررسی وجود توضیحات، محدودیت‌ها و قابلیت اتکای روش توضیح‌پذیری را نیز مورد توجه قرار دهد (بهاومیک و همکاران، ۲۰۲۲: ۲۰).

در لایه دوم، یعنی مستندسازی حاکمیتی، حسابرس بررسی می‌کند که آیا مسئولیت مدیریت ریسک مدل به صورت رسمی تعیین شده، فرایند بازبینی دوره‌ای مدل وجود دارد و سازوکار مشخصی برای شناسایی و مدیریت تغییرات عملکرد یا ریسک‌های مدل در نظر گرفته شده است. در این مرحله می‌توان از اصول حاکمیت و مدیریت کوبیت و الزامات سیستم مدیریت هوش مصنوعی ایزو ۴۲۰۰۱ به عنوان مبنای ارزیابی استفاده کرد (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی، ۲۰۲۵: ۱۳-۱۴؛ سازمان بین‌المللی استانداردسازی، ۲۰۲۳: ۲۰).

در لایه سوم، یعنی گزارشگری اطمینان بخشی، حسابرس یافته‌های مربوط به کمبود مستندات، محدودیت در قابلیت توضیح یا ضعف کنترل‌های حاکمیتی را در گزارش اطمینان بخشی یا حسابرسی فناوری اطلاعات منعکس می‌کند. در این پژوهش، این اقدام پیشنهادی مفهومی است، نه الزامی استاندارد.

بنابراین، چارچوب پیشنهادی تلاش می‌کند ارزیابی هوش مصنوعی را از یک موضوع صرفاً فنی فراتر برده و آن را به سه سطح مرتبط با یکدیگر شامل شفافیت فنی، حاکمیت سازمانی و گزارشگری اطمینان بخشی پیوند دهد. این چارچوب می‌تواند مبنایی مفهومی برای توسعه چک‌لیست‌های حسابرسی و انجام پژوهش‌های تجربی آینده در زمینه حسابرسی فناوری اطلاعات و حاکمیت هوش مصنوعی باشد.



بحث و نتیجه گیری

این پژوهش نشان داد که توضیح پذیری هوش مصنوعی، به عنوان بُعدی مستقل از عملکرد و دقت مدل، در ادبیات داخلی حسابرسی فناوری اطلاعات کمتر مورد توجه قرار گرفته، در حالی که ادبیات بین المللی و چارچوب های حرفه ای به طور فزاینده ای بر شفافیت، پاسخ گویی و حاکمیت سامانه های هوش مصنوعی تأکید می کنند (بهائومیک و همکاران، ۲۰۲۲: ۱۲؛ مک اینتاش و همکاران، ۲۰۲۴: ۱؛ انجمن حسابرسی و کنترل سیستم های اطلاعاتی، ۲۰۲۵: ۱۲) یافته های پژوهش باهدف نخست مطالعه همخوانی دارد و تا حد زیادی شکاف مطرح شده در مقدمه را تأیید می کند. مقایسه رویکردهای مطرح شده از سوی هیئت استانداردهای بین المللی حسابرسی و اطمینان بخشی با مطالعات انجام شده در داخل کشور نشان می دهد که میان این دو حوزه هنوز فاصله هایی وجود دارد و روند حرکت آن ها به سمت یک رویکرد مشترک و هماهنگ کامل نشده است (هیئت استانداردهای بین المللی حسابرسی و اطمینان بخشی، ۲۰۲۴ و ۲۰۲۶؛ سلیمانی امیری و محمدی سه دران، ۱۴۰۳: ۱). در چنین شرایطی، تدوین راهنماها و دستورالعمل های عملیاتی متناسب با ویژگی ها و الزامات محیط حسابرسی ایران می تواند به کاهش این فاصله و استفاده مؤثرتر از رویکردهای نوین در حسابرسی فناوری اطلاعات کمک کند.

ازسوی دیگر، نتایج پژوهش رونقی و محمودی (۱۳۹۴: ۶۱۵) درباره ارتباط میان حاکمیت فناوری اطلاعات و حاکمیت شرکتی نشان می دهد که موضوع حاکمیت هوش مصنوعی نیز نمی تواند صرفاً در محدوده وظایف واحد فناوری اطلاعات قرار گیرد و لازم است در سطح ساختار حاکمیتی سازمان نیز مورد توجه باشد. همچنین، باتوجه به گسترش استفاده از سامانه های هوش مصنوعی، افزایش دانش و مهارت حسابسان در زمینه مفاهیم پایه این فناوری و شیوه های ارزیابی توضیح پذیری مدل ها، می تواند نقش مهمی در تقویت توانمندی حرفه ای حسابرسی و پاسخ گویی آن به الزامات جدید در ایران داشته باشد.

مهم ترین نوآوری پژوهش حاضر، ارائه یک چارچوب مفهومی سه لایه شامل شفافیت فنی، مستندسازی حاکمیتی و گزارشگری اطمینان بخشی است. در این چارچوب، توضیح پذیری هوش مصنوعی صرفاً یک ویژگی فنی در نظر گرفته نشده، بلکه به عنوان یکی از اجزای مرتبط با حاکمیت و مدیریت ریسک فناوری اطلاعات و فرایند اطمینان بخشی حسابرسی مورد توجه



قرار گرفته است. همچنین، با در نظر گرفتن شرایط سازمانی ایران، این چارچوب می‌تواند مبنایی برای طراحی چک‌لیست‌ها و رویه‌های حسابرسی مرتبط با سامانه‌های هوش مصنوعی و نیز انجام مطالعات تجربی در آینده قرار گیرد.

نتایج کاربردی

یافته‌های پژوهش می‌تواند برای گروه‌های مختلفی از ذی‌نفعان حرفه‌ای و دانشگاهی قابل استفاده باشد. از جمله، مؤسسات حسابرسی می‌توانند با توجه بیشتری به مفاهیم مرتبط با هوش مصنوعی و روش‌های ارزیابی توضیح‌پذیری مدل‌ها بپردازند. با توجه به افزایش توجه نهادهای حرفه‌ای بین‌المللی به آثار فناوری‌های نوظهور بر فرایند حسابرسی و اطمینان‌بخشی، تقویت این مهارت‌ها می‌تواند به آماده‌تر شدن حساب‌برسان برای مواجهه با سامانه‌های هوش مصنوعی کمک کند (هیئت استانداردهای بین‌المللی حسابرسی و اطمینان‌بخشی، ۲۰۲۴ و ۲۰۲۶). همچنین، چارچوب سه‌لایه ارائه شده در این پژوهش می‌تواند نقطه شروعی برای تدوین رویه‌ها و چک‌لیست‌های مرتبط با حسابرسی فناوری اطلاعات در مؤسسات حسابرسی ایران باشد.

در مورد کمیته‌های حسابرسی، نتایج نشان می‌دهد که نظارت بر هوش مصنوعی نباید تنها در سطح واحد فناوری اطلاعات دنبال شود. موضوعاتی مانند مسئولیت‌پذیری، مدیریت ریسک مدل‌ها و نحوه مستندسازی تصمیمات الگوریتمی، نیازمند توجه در سطح مدیریتی و حاکمیتی سازمان هستند (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی، ۲۰۲۵: ۳؛ سازمان بین‌المللی استانداردسازی، ۲۰۲۳). از این رو، کمیته‌های حسابرسی می‌توانند در نظارت بر نحوه استفاده سازمان از سامانه‌های هوش مصنوعی و ارزیابی ریسک‌های مرتبط با آن نقش فعال‌تری داشته باشند.

برای نهادهای حرفه‌ای حسابرسی در ایران، از جمله انجمن حسابرسی فناوری اطلاعات ایران، یافته‌های این پژوهش می‌تواند زمینه‌ای برای تدوین راهنماها و دستورالعمل‌های اجرایی متناسب با شرایط کشور فراهم کند. استفاده از چارچوب‌هایی مانند کویت و ایزو ۴۲۰۰۱ در کنار در نظر گرفتن الزامات و شرایط محیط حرفه‌ای ایران، می‌تواند به شکل‌گیری رویه‌های منسجم‌تر در زمینه حسابرسی سامانه‌های هوش مصنوعی کمک کند.



همچنین، جدول ۲ پژوهش را می‌توان به‌عنوان یک ابزار اولیه برای بررسی وضعیت و آمادگی سازمان‌ها از منظر حاکمیت هوش مصنوعی در نظر گرفت. با این حال، از آنجاکه این جدول بر اساس چارچوب مفهومی پژوهش حاضر طراحی شده است، نمی‌توان آن را در وضعیت فعلی یک ابزار عملیاتی نهایی دانست. استفاده گسترده از آن در محیط حرفه‌ای، مستلزم آزمون و اعتبارسنجی تجربی و بررسی میزان قابلیت اتکای شاخص‌های آن در سازمان‌های مختلف است.

در مجموع، نتایج این پژوهش می‌تواند ارتباط میان مباحث نظری مربوط به توضیح‌پذیری هوش مصنوعی و نیازهای عملی حسابرسی فناوری اطلاعات را تا حدی روشن‌تر کند. چارچوب پیشنهادی نیز می‌تواند نقطه شروعی برای توسعه ابزارهای ارزیابی، تدوین رویه‌های حسابرسی و انجام مطالعات تجربی در زمینه حاکمیت و اطمینان بخشی سامانه‌های هوش مصنوعی در محیط سازمانی ایران باشد.

پژوهش حاضر از نظر روش‌شناختی ماهیتی مفهومی و مروری دارد و بر داده‌های حاصل از بررسی میدانی یا مطالعه موردی سازمان‌های مشخص استوار نیست. از این رو، چارچوب سه‌لایه ارائه‌شده در این پژوهش را نمی‌توان به‌عنوان چارچوبی نهایی و آزمون‌شده در محیط واقعی سازمانی تلقی کرد و بررسی میزان اعتبار و کارایی آن به پژوهش‌های تجربی آینده نیاز دارد. از سوی دیگر، باتوجه به جدید بودن استاندارد ایزو ۴۲۰۰۱ و شکل‌گیری تدریجی ادبیات مرتبط با سیستم‌های مدیریت هوش مصنوعی، هنوز شواهد کافی درباره نحوه استقرار، ارزیابی و کنترل الزامات این استاندارد در سازمان‌های ایرانی وجود ندارد؛ بنابراین، بخشی از چالش‌های اجرایی و الزامات بومی‌سازی آن همچنان نیازمند بررسی بیشتر است (مک‌این‌تاش و همکاران، ۲۰۲۴: ۱؛ سازمان بین‌المللی استانداردسازی، ۲۰۲۳).

محدودیت دیگر به تفاوت میان زمینه پژوهش حاضر و بخش قابل توجهی از مطالعات پیشین مربوط می‌شود. بسیاری از پژوهش‌های انجام‌شده در سایر کشورها، توضیح‌پذیری هوش مصنوعی و حسابرسی مدل‌ها را عمدتاً از منظر فنی بررسی کرده‌اند یا بر استفاده از هوش مصنوعی به‌عنوان ابزاری برای انجام فرایندهای حسابرسی تمرکز داشته‌اند. در مقابل، ارتباط میان توضیح‌پذیری مدل‌های هوش مصنوعی، حاکمیت و مدیریت ریسک فناوری اطلاعات و الزامات حسابرسی فناوری اطلاعات، به‌ویژه در شرایط نهادی، قانونی و سازمانی ایران، کمتر



مورد توجه قرار گرفته است. این موضوع نشان می‌دهد که برای دستیابی به چارچوبی متناسب با شرایط کشور، انجام مطالعات بومی و مبتنی بر تجربه واقعی سازمان‌ها ضروری است. با توجه به این محدودیت‌ها، چند مسیر برای پژوهش‌های آینده پیشنهاد می‌شود:

(۱) آزمون تجربی چارچوب پیشنهادی: چارچوب ارائه شده می‌تواند در یک یا چند سازمان ایرانی، به‌ویژه بانک‌ها و مؤسسات مالی، مورد مطالعه قرار گیرد. این سازمان‌ها به دلیل استفاده از سامانه‌هایی مانند اعتبارسنجی، تشخیص تقلب و تصمیم‌گیری الگوریتمی، بستر مناسبی برای بررسی عملی الزامات توضیح‌پذیری و حاکمیت هوش مصنوعی هستند.

(۲) بررسی دیدگاه حساب‌برسان: می‌توان با طراحی و اجرای پرسش‌نامه‌ای معتبر، میزان آمادگی و توانمندی حساب‌برسان داخلی و مستقل ایران برای ارزیابی حاکمیت هوش مصنوعی، مدیریت ریسک‌های مرتبط با آن و بررسی توضیح‌پذیری سامانه‌های هوش مصنوعی را مورد سنجش قرارداد.

(۳) مطالعه تطبیقی چارچوب‌های حاکمیتی: انجام پژوهش‌های تجربی برای مقایسه قابلیت‌ها و محدودیت‌های چارچوب‌هایی همچون کویت، ایزو ۴۲۰۰۱ و چارچوب مدیریت ریسک هوش مصنوعی در شرایط واقعی سازمان‌های ایرانی می‌تواند مشخص کند که هر یک از این چارچوب‌ها در کدام بخش از فرایند حاکمیت و حسابرسی هوش مصنوعی کاربرد بیشتری دارند (انجمن حسابرسی و کنترل سیستم‌های اطلاعاتی، ۲۰۲۵؛ سازمان بین‌المللی استانداردسازی، ۲۰۲۳).

(۴) تدوین ابزارهای اجرایی متناسب با شرایط ایران: همکاری میان نهادهای حرفه‌ای حسابرسی، از جمله انجمن حسابرسی فناوری اطلاعات ایران، دانشگاه‌ها و متخصصان حوزه فناوری اطلاعات می‌تواند به تدوین راهنماها، دستورالعمل‌ها و چک‌لیست‌هایی منجر شود که علاوه بر توجه به استانداردهای بین‌المللی، با الزامات قانونی، مقرراتی و نیازهای حرفه‌ای حساب‌برسان در ایران نیز سازگار باشند.

در مجموع، اجرای این پیشنهادها می‌تواند زمینه را برای بررسی تجربی چارچوب پیشنهادی و شناسایی نقاط قوت و ضعف آن در محیط واقعی فراهم کند. چنین مطالعاتی، علاوه بر کاهش فاصله میان مباحث نظری و الزامات عملی حسابرسی فناوری اطلاعات، می‌تواند به شکل‌گیری



ادبیات بومی در زمینه حاکمیت هوش مصنوعی، مدیریت ریسک مدل‌ها و حسابرسی سامانه‌های هوش مصنوعی در ایران کمک کند.

ملاحظات اخلاقی

حامی مالی: مقاله حامی مالی ندارد.

مشارکت نویسندگان: تمام نویسندگان در آماده‌سازی مقاله مشارکت داشته‌اند.

تعارض منافع: بنابر اظهار نویسندگان در این مقاله هیچ گونه تعارض منافی وجود ندارد.

تعهد کپی رایت: طبق تعهد نویسندگان حق کپی رایت رعایت شده‌است.

منابع

احمد حسین، سهیل؛ زواری رضایی، اکبر؛ آشتاب، علی. (۱۴۰۳). تحلیل اثرات حسابداری بخش عمومی، حاکمیت شرکی و کیفیت حسابرسی بر عملکرد مالی بانک‌ها قبل و پس از کووید-۱۹: رویکرد هوش مصنوعی. حسابداری و بودجه‌ریزی بخش عمومی، ۵(۴): ۷۴-۱۰۹.

<https://doi.org/10.22034/psab.2025.209761>

رونقی، محمدحسین؛ فرهادی، پیام؛ خردمند، صنوبر. (۱۳۹۷). ارزیابی حاکمیت فناوری اطلاعات با رویکرد سلسله‌مراتبی خاکستری (مورد مطالعاتی: دانشکده آموزش‌های الکترونیک دانشگاه شیراز). رشد فناوری، ۱۴(۵۴): ۹-۱۶.

<https://sid.ir/paper/145043/en>

رونقی، محمدحسین؛ محمودی، جعفر. (۱۳۹۴). رابطه حاکمیت فناوری اطلاعات و حاکمیت شرکی میان سازمان‌های دولتی حوزه فناوری اطلاعات. مدیریت فناوری اطلاعات، ۷(۳): ۶۱۵-۶۳۴.

<https://doi.org/10.22059/jitm.2015.54329>

سلیمانی امیری، غلامرضا؛ محمدی سه‌دران، سارا. (۱۴۰۳). حسابرسی در عصر هوش مصنوعی. حسابداری و منافع اجتماعی، ۱۴(۳): ۱-۳۶.

<https://doi.org/10.22051/jaasci.2024.47350.1862>

شاه صاحبی، سید مصطفی؛ دارابی، رؤیا؛ حمیدیان، محسن. (۱۳۹۹). تحلیل توانایی مدیریت با تأکید بر نقش شاخص‌های حسابداری و حاکمیت شرکی (روش هوش مصنوعی). دانش حسابداری و حسابرسی مدیریت، ۹(۳۳): ۷۵-۹۰.

<https://sid.ir/paper/392637/en>

صحت برمجه، اعظم؛ شماخی، حمیدرضا؛ یلفانی، امیر. (۱۴۰۵). ارائه الگوی قضاوت حسابرسی مبتنی بر هوش مصنوعی. دانش حسابداری و حسابرسی مدیریت، ۱۵(۶۰): ۱۹۱-۲۱۰.

<https://doi.org/10.22034/jmaak.2025.77853.4352>



References

- Bhaumik, D., Dey, D., & Kayal, S. (2022). A framework for auditing multilevel models using explainability methods. *The International Conference on AI Research Proceedings*, 4(1), 12–21. <https://doi.org/10.34190/icaire.4.1.874>
- Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273. <https://doi.org/10.3389/fhumd.2024.1421273>
- Hussein, S. A., Zavari Rezaei, A., & Ashtab, A. (2025). Analyzing the impact of public sector accounting, corporate governance, and audit quality on banks' financial performance before and after COVID-19: An artificial intelligence approach. *Journal of Public Sector Accounting and Budgeting*, 5(4), 74–109. <https://doi.org/10.22034/psab.2025.209761> (in Persian)
- IAASB. (2024). *Technology position statement: 8 guiding actions*. International Auditing and Assurance Standards Board. <https://www.iaasb.org/publications/technology-position-statement-8-guiding-actions>
- IAASB. (2026). *Technology quality management roundtables: Outcomes and next steps*. International Auditing and Assurance Standards Board. <https://www.iaasb.org/publications/technology-quality-management-roundtables-outcomes-and-next-steps>
- IIA. (2024). *Global internal audit standards*. The Institute of Internal Auditors.
- ISACA. (2025). *Leveraging COBIT for effective AI system governance*. ISACA. <https://www.isaca.org/resources/white-papers/2025/leveraging-cobit-for-effective-ai-system-governance>
- ISO. (2023). *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*. International Organization for Standardization. <https://www.iso.org/standard/81230.html>
- 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. [https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)
- McIntosh, T. R., Susnjak, T., Liu, T., Watters, P. A., Xu, D., Liu, D., Nowrozy, R., & Halgamuge, M. N. (2024). From COBIT to ISO 42001: Evaluating cybersecurity frameworks for opportunities, risks, and regulatory compliance in commercializing large language models. *Computers & Security*, 144, 103964. <https://doi.org/10.1016/j.cose.2024.103964>
- Ronaghi, M. H., Farhadi, P., & Kheradmand, S. (2018). Evaluating information technology governance by grey theory. *Roshd-e-Fanavari*, 14(54), 9–16. <https://sid.ir/paper/145043/en> (in Persian)
- Ronaghi, M. H., & Mahmoudi, J. (2015). The relationship between IT governance and corporate governance among IT-related government organizations. *Journal of Information Technology Management*, 7(3), 615–634. <https://doi.org/10.22059/jitm.2015.54329> (in Persian)
- Sehat Baramcheh, A., Shamakhi, H. R., & Yalfani, A. (2026). Providing an Audit Judgment Model based on Artificial Intelligence. *Journal of Management Accounting and Auditing Knowledge*, 15(60), 191–210. <https://doi.org/10.22034/jmaak.2025.77853.4352> (in Persian)



- Shahsahebi, S. M., Darabi, R., & Hamidian, M. (2021). Analysis of Ability Management with Emphasis on the Role of Accounting and Corporate Governance Indicators (Artificial Intelligence Method). *Journal of Accounting Knowledge and Management Auditing*, 9(33), 75–90. <https://sid.ir/paper/392637/en> (in Persian)
- Soleimany Amiri, G., & Mohammadi Sedaran, S. (2024). Auditing in the Age of Artificial Intelligence. *Journal of Accounting and Social Interests*, 14(3), 1–36. <https://doi.org/10.22051/jaasci.2024.47350.1862> (in Persian)

COPYRIGHT



This license allows others to download the works and share them with others as long as they credit them, but they can't change them in any way or use them commercially.

