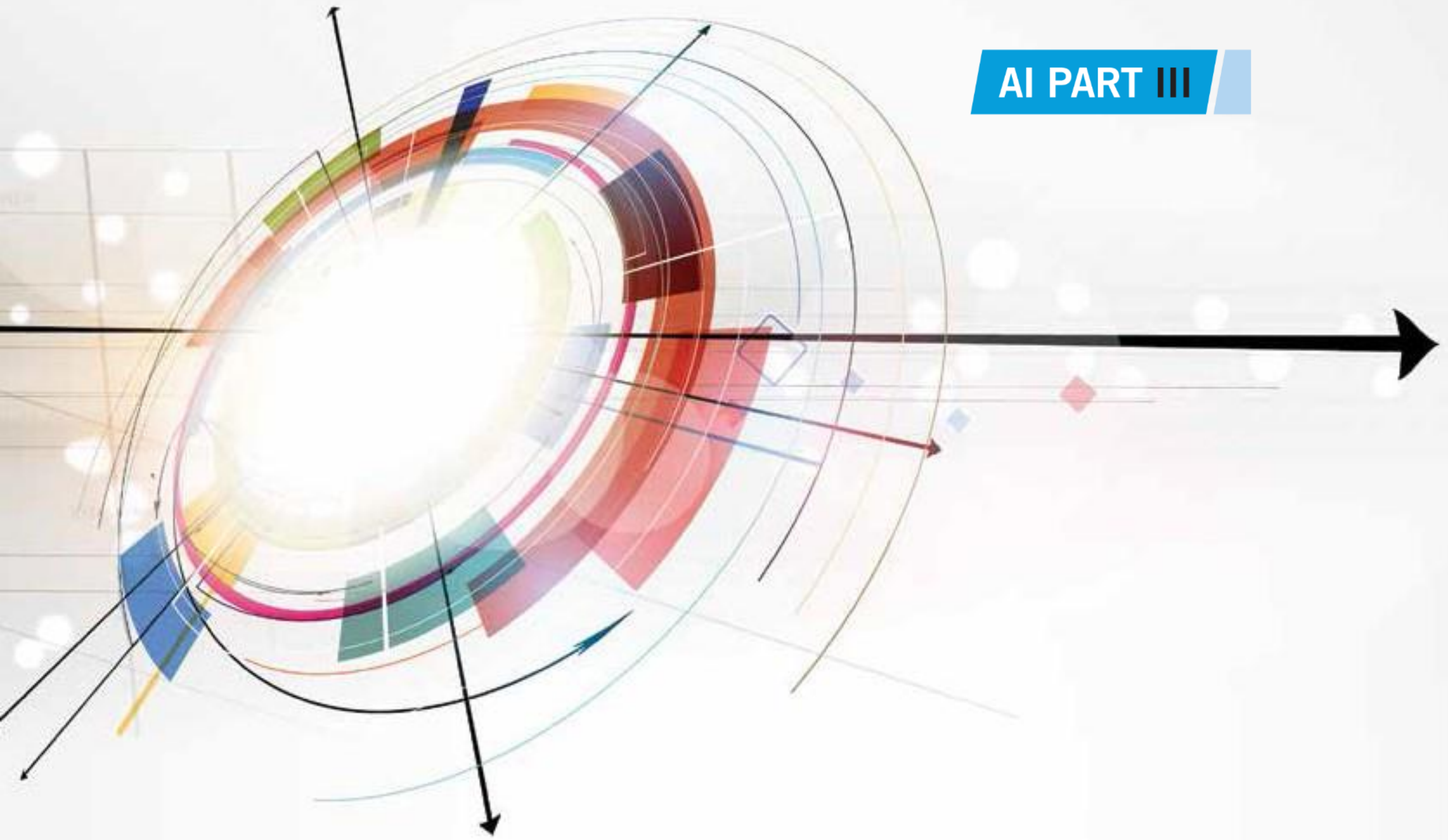
AI PART III



GLOBAL PERSPECTIVES AND INSIGHTS

The IIA's Artificial Intelligence Auditing Framework

Practical Applications, Part B

Special Edition

The Institute of Internal Auditors | Global

Advisory Council

Previous Issues

To access previous issues of Global Perspectives and Insights, visit www.theiia.org/gpi.

Reader Feedback

Send questions or comments to globalperspectives@theiia.org.

Table of Contents

Introduction

Picture it. A robot — or least its brain — before an *ethics* or investigative committee. It could happen. According to The Guardian, academics argue that as robots start to enter public spaces, and work alongside humans, the need for safety measures has become more pressing. Scientists are attempting to make a case for robots to be fitted with an "ethical black box" to keep track of their decisions, and enable them to explain — yes, explain — their actions when *accidents* happen.

Professors at Oxford University argue that robotics firms should follow the examples and regulations set by the aviation industry, which include bringing in the black boxes and cockpit voice recorders to investigate plane crashes, ensuring that crucial safety lessons are learned after such tragic events. Installed in a robot, an ethical black box would record the robot's decisions, its basis for making them, its movements, and information from sensors such as cameras, microphones, and rangefinders.

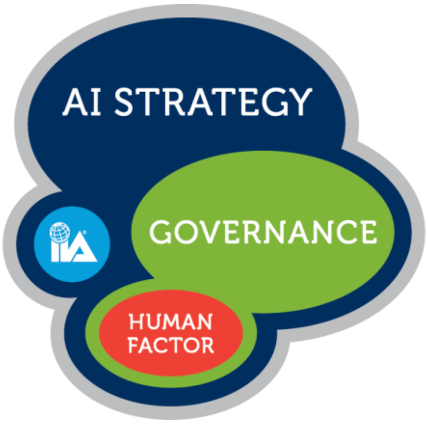
These actions come in the wake of recent incidents, such as when "Steve," a Knightscope K5 patrol robot, fell down steps and plunged into a fountain while on duty, or when another K5 robot became involved in a carpark altercation with a 41-year-old man while patrolling the streets of Mountain View, Calif., or the robotic unit accused of running over a 16-month-old child in a Stanford shopping center. All injuries of humans involved were minor.

There are hundreds of reports on the ethics of artificial intelligence; however, most are *lightweight* and full of platitudes about putting people first, writes Scott Rosenberg, editor of *Backchannel*. Rosenberg extracts from recently released reports from New York University's AI Now Institute about a tech industry attempting to reshape society along AI lines without any guarantee of reliable and fair results. One report concludes that "efforts to hold AI to ethical standards to date, have been a flop, and that new ethical frameworks for AI need to move beyond individual responsibility to hold powerful industrial, governmental, and military interests accountable as they design and employ AI." In authors' opinions, AI systems are being introduced in various and vulnerable areas, such as policing, education, healthcare, and other environments, where the "misfiring of an algorithm could ruin a life."

The IIA's AI Auditing Framework

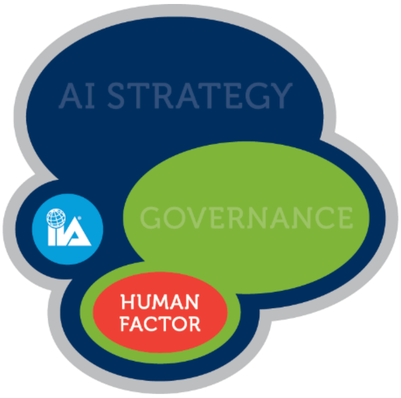
As explained in Artificial Intelligence – Considerations for the Profession of Internal Auditing, internal audit's role in AI is to "help an organization evaluate, understand, and communicate the degree to which artificial intelligence will have an effect (negative or positive) on the organization's ability to create value in the short, medium, or long term."

Note: This is the third report in a three-part series. For more information, see Artificial Intelligence – Considerations for the Profession of Internal Auditing and The IIA's Artificial Intelligence Auditing Framework: Practical Applications Part A.



To help internal audit fulfill this role, internal auditors can leverage The IIA's AI Auditing Framework in providing AI-related advisory, assurance, or blended advisory/assurance services as appropriate to the organization. The Framework is comprised of three overarching components — AI Strategy, Governance, and the Human Factor — and seven elements: Cyber Resilience; AI Competencies; Data Quality; Data Architecture & Infrastructure; Measuring Performance; Ethics; and The Black Box.

Internal audit should consider numerous engagement or control objectives, and activities or procedures, in implementing the Framework and providing advisory, assurance, or blended advisory/assurance internal audit services related to the organization's AI activities. Relevant objectives and activities or procedures that address the Strategy (Cyber Resilience and AI Competencies elements) and Governance (Data Architecture & Infrastructure, and Data Quality elements) of the Framework were provided in The IIA's Artificial Intelligence Auditing Framework: Practical Applications Part A. This document provides relevant objectives and activities or procedures that address the Human Factor (Ethics and The Black Box elements) and Governance (Measuring Performance element).



The Human Factor

The Human Factor component, which includes Ethics and The Black Box elements, addresses the risk of human error compromising the ability of AI to deliver the expected results.

Ethics

Algorithms developed by humans that include human error and biases (both intentional and unintentional) will impact the performance of the algorithm. The human factor component considers whether:

- The risk of unintended human biases factored into AI design is identified and managed.
- AI has been effectively tested to ensure that results reflect the original objective.
- AI technologies can be transparent given the complexity involved.
- AI output is being used legally, ethically, and responsibly.

Algorithm Bias

According to a recent McKinsey & Company report, companies are quick to apply machine learning to business decision-making. The programs set complex algorithms to work on large, frequently refreshed data sets. However, algorithmic bias is risky business, because it can compromise the very purpose of machine learning if overlooked, and left unchecked (see Controlling machine-learning algorithms and their biases).

For example, in credit scoring, the customer with a long history of maintaining loans without delinquency or default is generally determined as "low risk." However, what may be unseen is that the mortgages of this customer have been maintained and supported by substantial tax benefits that are set to expire. A spike in defaults may be in the offing, unaccounted for in the statistical risk model of the lending institution. With access to the right data and guidance by subject matter experts, predictive machine-learning models could find the hidden patterns in the data and correct for such spikes. Even more, outside of *business* decisions, algorithmic bias can cause errors that could spark some real problems and unrest among citizens. For example, Google's Photos service and others services like it are used to identify people, objects, and scenes, but can go terribly wrong, such as when a camera missed the mark on racial sensitivity, or when a software used for risk assessments to predict future criminals showed bias.

Meaning Making

There are limits to what machines can do, and *people* must be able to make sense of AI outputs. As described in a report by McKinsey & Company, "Meaning making in the AI era starts with an appreciation of what machines can and cannot do. It may be possible, for example, for a machine to make certain kinds of diagnoses more accurately than a person can. But it will be up to nurses, doctors, and therapists to help patients understand the implications and manage the consequences. It's the difference between knowledge and meaning." According to the report, hard skills such as coding, analytics, and data science are critical to AI, but so are soft skills such as collaboration, empathy, and meaning making (see McKinsey Quarterly pp. 56-61).

Relevant Ethics Objectives and Activities or Procedures

Engagement or Control Objective (s)	Activities or Procedures
Provide assurance that outcomes of the organization's AI activities are free from unintended biases.	Review the intended results of the AI activities (strategic objectives) and compare with actual results. If a variance is detected, determine if bias is the cause.
The organization can "make meaning" of AI outputs.	Review AI outputs and the meaning that was derived from the outputs.

The Black Box

According to the *Merriam-Webster* online dictionary, a black box is "a usually complicated electronic device whose internal mechanism is usually hidden from or mysterious to the user; *broadly*: anything that has mysterious or unknown internal functions or mechanisms." As organizations advance to implementing Type III and Type IV AI technologies — utilizing machines or platforms that can learn on their own or communicate with each other — how the algorithms are operating becomes less transparent or understandable.

Relevant objectives and activities or procedures identified by The IIA do not comprise a prescribed audit plan, but are examples that should be useful in identifying engagement or control objectives, and in planning and performing AI audit engagements.

AI audit engagements should conform with IIA Standard 2200: Engagement Planning. AI audit plans and IA engagement objectives and procedures should always be customized to meet the needs of the organization.

The black box factor will become more and more of a challenge as an organization's AI activities become more sophisticated.

Relevant Black Box Objectives and Activities or Procedures	
Engagement or Control Objective (s)	**Activities or Procedures**
Assess the organization's understanding of "black box" data (i.e., the underlying algorithms, internal functions, or mechanisms that enable AI).	**Review** AI development and implementation policies, processes, and procedures and verify black box data has been identified. **Interview** those responsible for AI outcomes and **verify** that they understand and could explain black box data.

AI STRATEGY
GOVERNANCE
HUMAN FACTOR

Governance

AI governance refers to the structures, processes, and procedures implemented to direct, manage, and monitor the AI activities of the organization. Governance structure and formality will vary based on specific characteristics of the organization. AI governance:

- Establishes accountability, responsibility, and oversight.
- Helps to ensure that those with AI responsibilities have the necessary skills and expertise.
- Helps to ensure that AI activities and AI-related decisions and action are consistent with the organization's values, ethical, social, and legal responsibilities.

Relative to the Governance component, Part II of this series, The IIA's Artificial Intelligence Auditing Framework: Practical Applications Part A, addressed accountability, responsibility, and oversight; regulatory compliance; data architecture and infrastructure; and data quality. This document addresses measuring performance.

Measuring Performance

Internal audit is positioned to become vital to the organization's ability to measure performance of its AI initiatives. In the planning stage, internal audit can provide advice on how to establish metrics that provide management and the board with sufficient, reliable, relevant, and useful information. However, internal audit must not be responsible for establishing or own AI performance metrics. In organizations where AI has been implemented, internal audit should provide assurance over first line of defense controls and second line of defense oversight related to AI. There is perhaps no better way to demonstrate internal audit's competence in AI than by using AI technologies, such as robotic process automation to audit AI.

Relevant Measuring Performance Objectives and Activities or Procedures

Engagement or Control Objective (s)	Activities or Procedures
Provide advice on how to establish AI metrics.	**Facilitate** working sessions for those responsible to establish AI metrics. **Convey** the importance and meaning of the terms *sufficient, reliable, relevant, and useful* information.
Stress test AI vulnerabilities.	**Apply** stress-testing techniques used by the banking industry to determine how AI activities will perform under extreme scenarios.
Communicate the results of AI-related engagements.	**Communicate** the results of AI-related engagements in conformance with: ▪ IIA Standard 2400: Communicating Results. ▪ IIA Standard 2410: Criteria for Communicating. ▪ IIA Standard 2420: Quality of Communications. ▪ IIA Standards 2421: Errors and Omissions. ▪ IIA Standard 2430: Use of "Conducted in Conformance with the *International Standards for the Professional Practice of Internal Auditing.*" ▪ IIA Standard 2431: Engagement Disclosure of Nonconformance. ▪ IIA Standard 2440: Disseminating Results.
Provide assurance over first line of defense controls and second line of defense oversight related to AI.	**Embrace** robotics and other forms of AI to perform AI-related engagements.

Closing Thoughts

The internal audit profession cannot be left behind in what may be the next digital frontier — artificial intelligence. To prepare, internal auditors must understand AI basics, the roles that internal audit can and should play, and AI risks and opportunities.

Whether the organization's AI technologies are developed in-house, through a facilitative technology, or by a third party, internal audit should be prepared to advise the board and senior management, coordinate with the first and second lines of defense, and provide assurance over AI risk management, governance, and controls. To meet these challenges, internal auditors should leverage The IIA's Artificial Intelligence Framework.

About The IIA

The Institute of Internal Auditors (IIA) is the internal audit profession's most widely recognized advocate, educator, and provider of standards, guidance, and certifications. Established in 1941, The IIA today serves more than 190,000 members from more than 170 countries and territories. The association's global headquarters are in Lake Mary, Fla., USA. For more information, visit www.globaliia.org.

Disclaimer

The opinions expressed in Global Perspectives and Insights are not necessarily those of the individual contributors or of the contributors' employers.

Copyright

The Institute of Internal Auditors | Global

globaliia.org