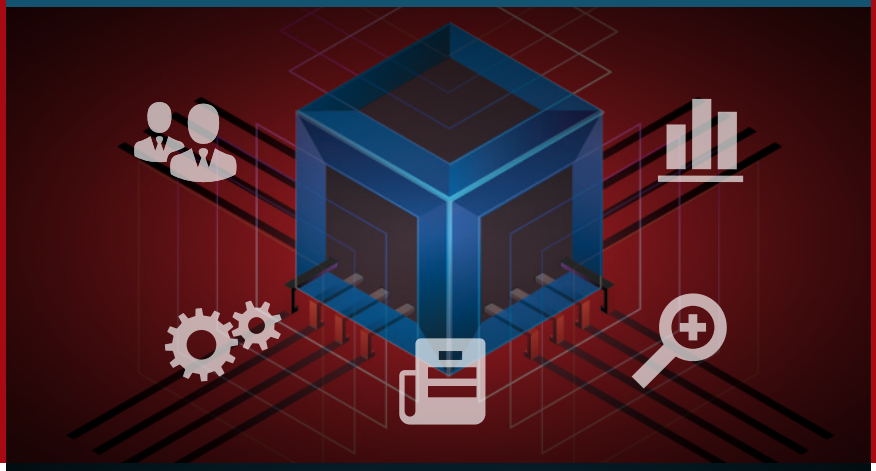
COSO

Committee of Sponsoring Organizations of the Treadway Commission

Enterprise Risk Management



REALIZE THE FULL POTENTIAL OF ARTIFICIAL INTELLIGENCE

APPLYING THE COSO FRAMEWORK AND PRINCIPLES TO HELP IMPLEMENT AND SCALE ARTIFICIAL INTELLIGENCE

Sponsored By

Deloitte.

Keri Calagna | Brian Cassidy | Amy Park

September 2021

Authors

Keri Calagna
Risk & Financial Advisory Principal
Deloitte & Touche LLP

Brian Cassidy
Audit & Assurance Partner
Deloitte & Touche LLP

Amy Park
Audit & Assurance Partner
Deloitte & Touche LLP

Acknowledgements

We would like to recognize and thank John Fogarty, Senior Manager, Deloitte & Touche LLP, Hemant Dhengane, Manager, Deloitte & Touche LLP, Mary Schmidlin, Senior Manager, Deloitte & Touche LLP, and Edward Bowen, Managing Director, Deloitte & Touche LLP for their technical input and advice.

The COSO Board would like to thank Deloitte & Touche LLP for its support.

COSO Board Members

Preface

This project was commissioned by the Committee of Sponsoring Organizations of the Treadway Commission (COSO), which is dedicated to helping organizations improve performance by developing thought leadership that enhances internal control, risk management, governance, and fraud deterrence.
COSO is a private-sector initiative jointly sponsored and funded by the following organizations:

American Accounting Association (AAA)

American Institute of CPAs (AICPA)

Financial Executives International (FEI)

The Institute of Management Accountants (IMA)

The Institute of Internal Auditors (IIA)

COSO
Committee of Sponsoring Organizations of the Treadway Commission

coso.org

APPLYING THE COSO FRAMEWORK AND PRINCIPLES TO HELP IMPLEMENT AND SCALE ARTIFICIAL INTELLIGENCE

Research Commissioned by

COSO

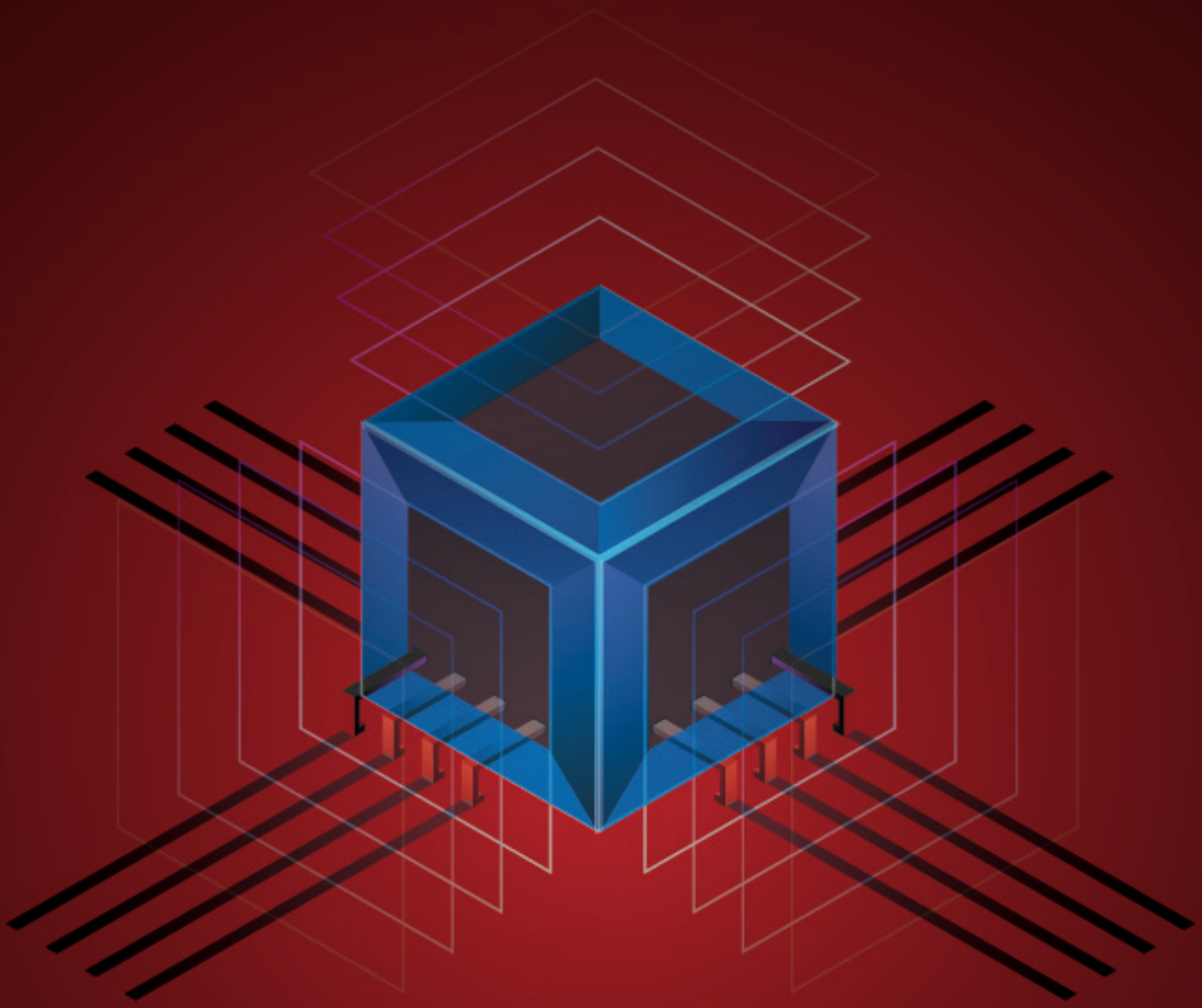
Committee of Sponsoring Organizations of the Treadway Commission

September 2021

1234567890 PIP 198765432

Design and production: Sergio Analco.

Contents



INTRODUCTION

Artificial intelligence (AI) has and will continue to transform business strategies, solutions, and operations. AI-related risks need to be top of mind and a key priority for organizations to adopt and scale AI applications and to fully realize the potential of AI. Applying enterprise risk management (ERM) principles to AI initiatives can help organizations provide integrated governance of AI, manage risks, and drive performance to maximize achievement of strategic goals. The COSO ERM Framework, with its five components and twenty principles, provides an overarching and comprehensive framework, can align risk management with AI strategy and performance to help realize AI's potential.

Figure 1. **COSO *Enterprise Risk Management – Integrating with Strategy and Performance Framework***

2017 COSO Enterprise Risk Management – Integrating with Strategy and Performance

THE AI REVOLUTION: TRANSFORMING BUSINESS AND INNOVATION

As AI expands into almost every aspect of modern life, it's becoming a required business capability. Whether it's managing customer relationships, identifying and responding to cyber threats, or helping guide medical decisions, AI is addressing a wide range of business issues. The rapid adoption of AI is providing insight into organizations' data that, in turn, provides intelligence to support decision-making. This has led to organizations investing in AI initiatives at a massive scale. AI spending is forecast to double by 2024, growing from $50.1B in 2020 to over $110B in 2024. The forecasted compound annual growth rate (CAGR) for this period is approximately 20%.[1] Furthermore, worldwide revenues for the AI market, including software, hardware, and services, are forecast to grow to $327.5B in 2021 and reach $554.3B by 2024 with a five-year CAGR of 17.5%.[2]

What's fueling the revolution? Organizations are applying AI for its transformative potential: to automate business processes, tasks, and actions to reduce costs, increase efficiency, and improve predictability of outcomes. With AI, they are seeing better data insights, leading to more informed business decisions, positive business and operational results, and increased innovation.

How organizations are using AI to drive value

COST REDUCTION
Applying AI to intelligently automate business processes, tasks, and interactions to reduce cost, increase efficiency, and improve predictability.

SPEED TO EXECUTION
Applying AI to accelerate time to operational and business results by minimizing latency.

PREDICTIVE ANALYTICS
Applying AI to provide insight into an organization's data and to improve understanding and decision-making by deciphering patterns, connecting dots, and predicting outcomes from increasingly complex data sources.

DIGITAL ENGAGEMENT
Applying AI to change how humans interact with smart systems by expanding the means of engagement via voice, vision, text, and touch.

FUELED INNOVATION
Applying AI to generate insights for new products, market opportunities, and business models.

Recent studies indicate that organizations are moving to take advantage of these benefits with near-term investments in AI:

- 75% of respondents expect to shift from piloting to operationalizing AI by the end of 2024.[3]
- 75% of surveyed AI adopters are expecting organizational transformation within three years.[4]
- 61% of surveyed AI adopters are anticipating industry transformation within the same timeframe.[5]
- Surveyed AI adopters are investing significantly, with 53% spending more than $20 million in 2020 on AI-related technology and talent.[6]
- 71% of surveyed AI adopters expect to increase investment in the next fiscal year, by an average of 26%.[7]

[1] International Data Corporation (IDC), "Worldwide Spending on Artificial Intelligence is Expected to Double in Four Years, Reaching $110 Billion in 2024, According to New IDC Spending Guide," August 25, 2020. https://www.idc.com/getdoc.jsp?containerId=prUS46794720

[2] International Data Corporation (IDC), "IDC Forecasts Improved Growth for Global AI Market in 2021," February 23, 2021. https://www.idc.com/getdoc.jsp?containerId=prUS47482321

[3] Gartner, Accelerating AI Deployments – Paths of Least Resistance, July 2020.

[4] Deloitte, State of AI in the Enterprise, 3rd Edition, 2020. Figure 2, page 7.

[5] Ibid., Figure 2, page 7.

[6] Ibid., page 6.

[7] Ibid., page 6.

To put organizational and industry transformation in perspective, many companies are investing in AI capabilities to pivot their business strategy. In some cases, AI underpins business models, such as the case of some financial technology companies moving away from traditional FICO scores and using multiple AI-powered parameters and models to inform credit decisions. The process is automated, making the effort more efficient, and it alerts users when cases need further review. It may improve decision-making and can enhance existing services and experience for customers.

AI and Machine Learning: A practical introduction

An understanding of AI-associated algorithms and how they're built is imperative to properly identify and manage AI-related risk. In practice, AI is developed by humans through the use of software programming (code). Similar to needing governance and controls in financial reporting or software development, due to the human element, organizations need governance and controls for AI as well. But boards and executives can't effectively help monitor controls without a basic understanding of what AI does and how it is built.

What algorithms do

There are three common classes of machine learning algorithms: non–deep-learning, deep-learning, and reinforcement learning. The goal of these AI models is to create a classification, a prediction, or the generation of novel data.

- **Non–deep-learning classifies, finds patterns, and predicts outcomes.** Common models include regressions, clustering, decision trees, and support vector machines. They can help with many useful and common problems such as demand forecasting, cross-selling propensity, and risk classification.
- **Deep-learning algorithms have been a game changer.** These methods of classifying and predicting have driven the AI revolution of the last decade. Imaging, natural language processing, and anomaly detection have achieved state-of-the-art results using deep neural networks. The conversational bots that are helping people navigate customer service on a website comes from this AI technology. A simple automation can be applied more widely, such as voice-to-text on a cell phone, or it can be used to recognize and translate handwriting, utilizing the data to aid in the effort.
- **Reinforcement learning models** examine an environment and develop the ability to make a sequence of decisions that aims to find the best positive path forward. Such models can learn to win Chess and Go tournaments against human grandmasters. Practical applications include route optimization, factory optimization, and cyber vulnerability testing.

How algorithms are built

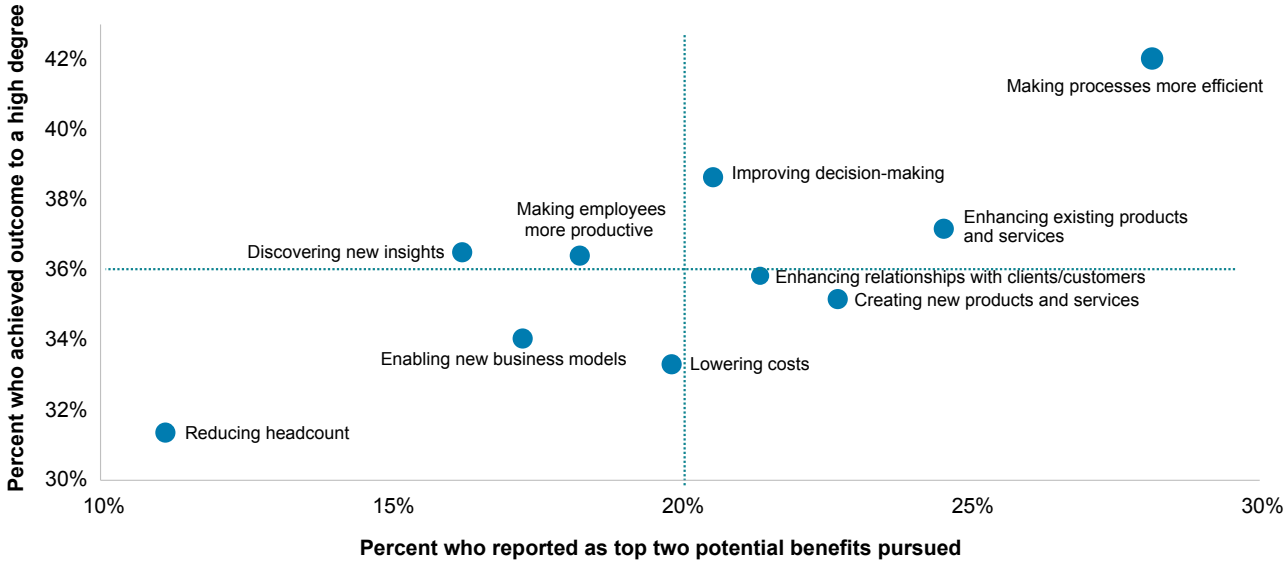
Every algorithm should link to the business strategy. Algorithms are designed by humans to contribute to informed decision-making that creates the intended business value. There are six key steps to building a machine learning model:

1. **Problem definition –** Considering a business problem and how machine learning could solve it.
2. **Data profiling –** Identifying the data sources needed to solve the problem and what additional data is needed. An emerging trend within AI is the development of new sensors and data collection for the sole purpose of improving AI performance. Organizations need to ensure that data is fair and balanced across ethical and performance dimensions.
3. **Data preparation –** Determining what's needed to transform, normalize, and cleanse the data, and creating a testing and validation approach.
4. **Algorithm evaluation –** Leveraging leading practices to select the algorithms required to solve the problem. Often, data science teams will develop multiple algorithms in parallel to determine the best performing model. It's important to establish the correct performance evaluation criteria.
5. **Model development –** Training, testing, and validating all identified algorithms with the data and implementing approaches like regularization.
6. **Model deployment, monitoring, and maintenance –** Incorporating machine learning operations (MLOps) and monitoring structures along with processes to address model drift. Model performance can degrade if the activities in the environment change over time (for example, models that predict electricity consumption need to be updated over time as solar panels gain traction with consumers).

AI serves a wonderful world … until there's an unfortunate outcome

As AI and machine learning deployment has increased, the top two benefits of deployment cited by surveyed adopters are increased process efficiency and enhancement of existing products and services. (See Figure 2) In addition, a survey conducted by Gartner indicates that the top two reasons for organizations to invest in AI capabilities are a desire to achieve an increase in revenue or a reduction in costs, and addressing vulnerabilities from competitors and start-ups.[8]

Figure 2. **Process Efficiency Tops the List of Benefits achieved with AI**



Blue dotted lines represent the average of respective dimensions

Source: *State of AI in the Enterprise, 3rd Edition*, Deloitte

AI drives efficiency through computer algorithms that use data to build predictions or prescriptive recommendations, generate classifications, and invent novel constructs. Many AI use cases implemented today are doing things humans can do but doing them much faster and more efficiently. Over the next ten years, the emphasis will likely evolve to implementing AI to do things humans can't do because humans are unable to see the subtlety and nuances that AI can detect. For example, pharmaceutical companies can use AI to interpret nuances in microscopic images that human scientists can't detect. This large-scale image-based cell profiling is quickly ascertaining the differences between large data sets of healthy and diseased cells in order to design highly specific new drug compounds to treat disease. In theory, researchers could make the comparisons by eye; however, comparing thousands of cells with tiny but consistent differences would be very difficult without the use of AI. In essence, AI is driving transformative innovation. These trends may further accelerate or evolve in the future.

Although AI seems like a panacea for business transformation, the technology and application of the technology is not without risks that could result in serious problems for an organization. Those risks can be mitigated by thoughtful and pre-emptive consideration of the COSO ERM Framework. But first, let's talk about the risks. There is a broad spectrum of AI-related risks that include, but are not limited to the following:

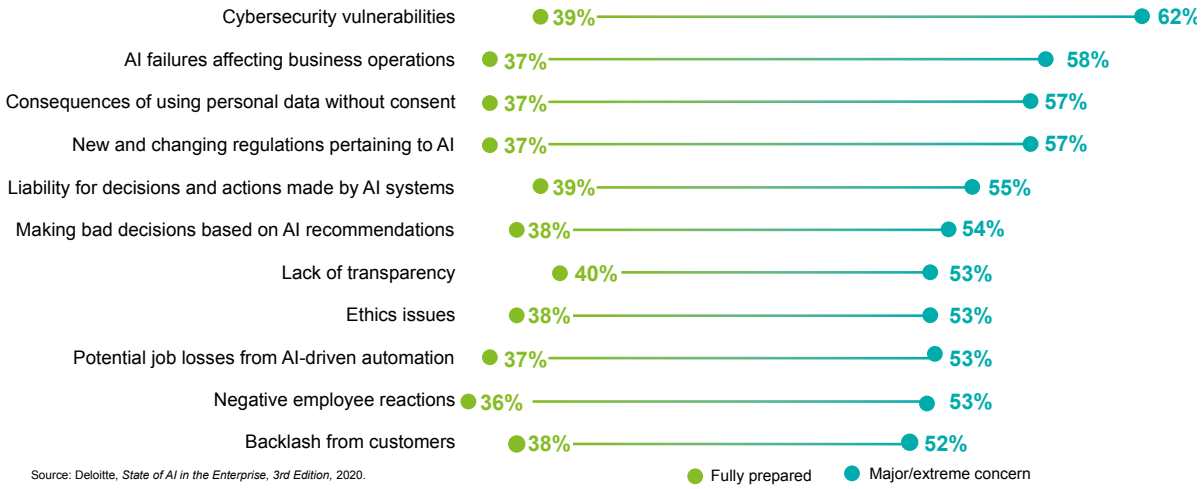
- Bias and reliability breakdowns due to inappropriate or non-representative data
- Inability to understand or explain AI model outputs
- Inappropriate use of data
- Vulnerabilities to adversarial attack to obtain data or otherwise manipulate the AI model
- Societal stresses due to rapid application and transformation of AI technologies

[8] 2019 Gartner, AI in Organizations Survey. 735439_C.

Potential consequences from these risks can include reputational damage, destruction of shareholder value, regulatory fines, and lawsuits. Because of such emerging risks, 56% of surveyed AI adopters say their organization is slowing the adoption of AI technologies.[9] However, that may not be feasible for long if organizations are going to remain competitive. Rather than tapping the brakes, a more prudent strategy may be to better manage associated risks. Organizations cannot ignore risks or unintended consequences of AI.

Deloitte's "State of AI in the Enterprise" survey illustrates that AI implementers and adopters have serious concerns about the use of AI that span a variety of risk areas beyond bias. (See Figure 3) Furthermore, respondents to the survey indicate that there are significant gaps in their organizations' current abilities to address these concerns. Results from a separate survey conducted by Gartner cited the top barriers to AI implementation as security or privacy concerns and complexity of AI solution(s) integration with existing infrastructure.[10]

Figure 3. **Comparison of concern vs. preparedness for AI-related risks**



Source: Deloitte, *State of AI in the Enterprise, 3rd Edition,* 2020.

Impact of regulatory uncertainty

Regulatory requirements are another important consideration and adhering to regulatory compliance means not only following today's legislation, but also demonstrating commitment to safe AI practices that may become required in the future. Organizations should consider the applicable extent of pending regulatory requirements in evaluating their governance framework over AI and related data.

Figure 4. **Regulatory compliance**

Example Players

- The World Economic Forum's Council on the Future of AI and Robotics
- AI Now Initiative
- The Stanford One Hundred Year Study on AI
- MIT Media Lab, AI, Ethics and Governance Project
- The Partnership on AI
- Data & Society's Intelligence and Autonomy Initiative
- The IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems
- The Cambridge Center for the Study of Existential Risk

Example Standards, Policy and Laws

Example Regulation	What it Means for Your Business
Product liability laws apply to individuals injured when using an AI-driven product	Companies must monitor AI in the same way they do human employees (digital does not equal infallible)
The Restatement of Torts relates to AI design and manufacturing defects, and failure to warn	Extra care must be taken in developing warning labels for AI-driven products, e.g. "This product was audited with AI"
Fair Credit Reporting Act, and the FTC's enforcement against AI collusion	Conclusion could be unintentional without transparency into AI methods, meaning companies need strong control in place
EU General Data Protection Regulation affecting US companies operating in EU	Companies need to design policies around AI that meet expectations in even the most highly regulated markets
Speech laws applied to communication between bots and people	Extra controls must be implemented around conversational AI use cases to incorporate speech laws
Bot Disclosure and Accountability Act of 2018 to regulate news bots	Social media bots already require disclosure that they are operating on AI; future regulation may go beyond social bots

[9] Ibid., page 13.

[10] 2019 Gartner, AI in Organizations Survey. 729419_C.

THE COSO ERM FRAMEWORK: ADDRESSING AI RISKS ALIGNED WITH YOUR OVERALL BUSINESS AND IT STRATEGY

As AI becomes more pervasive in business and our everyday lives, organizations will likely no longer have the option of ignoring or avoiding the unique risks that accompany AI adoption. Instead, they must learn to identify and manage these risks effectively. Compounding the problem is the fact that AI is often not isolated to a specific function such as IT, but rather affects multiple functions in an organization. Organizations need to design and implement governance, risk management, and control strategies and structures to realize the potential of humans collaborating with AI. Fortunately, AI is like other technological components of an organization and thus can be successfully governed by effective ERM.

Since 1985, the voluntary, private-sector Committee of Sponsoring Organizations of the Treadway Commission (COSO) has been focused on helping organizations improve performance by developing thought leadership that enhances internal control, risk management, governance and fraud deterrence. The most recent update of the COSO ERM Framework – adopted in 2017 – highlights the importance of embedding it throughout an organization in five critical components:

- **Governance & Culture**
- **Strategy & Objective-Setting**
- **Performance**
- **Review & Revision**
- **Information, Communication, & Reporting**

Figure 5. **COSO *Enterprise Risk Management – Integrating with Strategy and Performance Framework***

2017 *COSO Enterprise Risk Management – Integrating with Strategy and Performance*

By leveraging the COSO ERM Framework, organizations can identify and manage AI-specific risks and establish practices to optimize the results while managing exposure to risks like unintended bias and lack of transparency. Implementation can help to improve confidence among stakeholders within and outside the organization, and proactively address emerging risks related to AI.

GOVERNANCE & CULTURE

Governance and culture together form the basis for all risk management components. Governance reinforces the importance of ERM and culture is reflected in decision-making at all levels within an organization. According to the COSO ERM Framework, these components must incorporate an organization's commitment to its vision, mission, and core values. Core values provide an important foundation for appropriate oversight of AI initiatives and AI models to help achieve the organization's strategy and business objectives. The Governance & Culture component and the following principles of the COSO ERM Framework serve as the basis for this section of the paper:

1. **Exercises board risk oversight**
2. **Establishes operating structures**
3. **Defines desired culture**
4. **Demonstrates commitment to core value**
5. **Attracts, develops and retains capable individuals**

An organization's board is often not involved in AI initiatives, or may not be fully aware of them to ask the appropriate risk-related questions of management. When high-level executives and board members understand AI and its implications and are actively engaged, they set the tone from the top about the importance of risk management. Such engagement is imperative.

Only about 26% of surveyed AI adopters have a single executive responsible for managing AI-related risks.[11] Similar to other core elements of a business, board members need to understand an organization's framework for evaluating risk associated with AI initiatives and determine the threshold of risk that requires oversight from senior leadership. Some initiatives may be limited to a small number of simple AI models and have a lower risk profile. Other initiatives may have a large number of complex AI models or touch critical business activities like delivering patient health care, ensuring customer safety, or controlling manufacturing activities and have a higher risk profile. High-risk AI initiatives require close oversight by a senior executive, who collaborates with a chief risk officer or equivalent risk leader. Organizations may need to acquire personnel with expertise in AI development and data analysis to properly oversee their AI initiatives or seek external advisers with the relevant experience if the needed skillset is missing at the organization. These individuals can advise board members, provide insights into risks/rewards and promote risk-informed decision-making. Such involvement is critical to effective adoption and implementation of AI and prevention of organizational crisis events.

The Importance of Governance

As AI is implemented on a broader scale within organizations, governance has a key role in appropriate oversight of AI initiatives and related models. Organizations are facing increased scrutiny from various stakeholders (e.g., regulators, customers, users, etc.) due, in part, to perceived inadequate oversight of AI.

Governance plays a key role in the following key areas:

1. To support the development and operation of AI models, organizations are collecting unprecedented amounts of data. Participants have concerns, including but not limited to, how their data is being used and who else has access to their data. Organizations need to have clear rules regarding use of data, collection of data, retention of data, and access of data and consistently apply those rules throughout the organization as part of their response to those concerns. Failure to appropriately address these issues can harm people and inflict damage on corporate reputation and shareholder value.
2. Organizations are increasingly applying AI to situations that require more judgment and may have a significant impact on participants. AI models that perform or inform significant judgments (e.g., underwriting decisions, eligibility for various benefits, medical diagnosis, and recommended treatment, etc.) that have a significant impact on participants may introduce ethical concerns. As part of their response, organizations need to assess when, where, and how AI is or will be used and whether such use is consistent with the organization's values and design, and how the organization's oversight structures engage with larger societal concerns, if applicable.

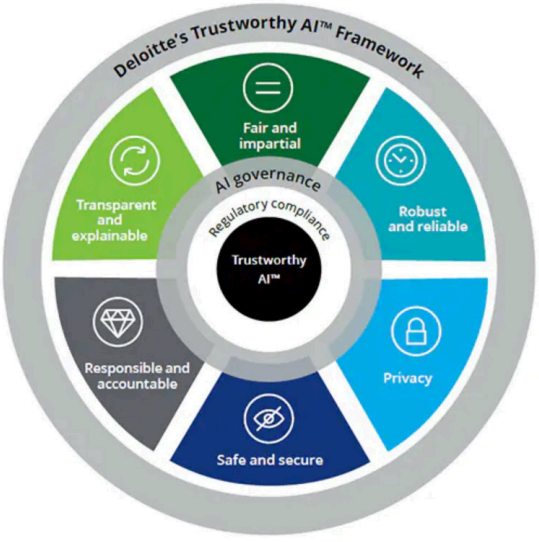
[11] Ibid., based on average from Figure 9 on page 15.

In addition, leaders need to understand how they define success when developing, deploying, monitoring, and maintaining AI and how it correlates to their company's purpose. Important aspects of defining success include determining which measures or metrics are most applicable as well as how the organization identifies and assesses costs versus benefits. Those aspects are closely related to management tying AI initiatives with the organization's broader commitment to its core values by providing the basis for enforcing accountability for actions and aligning risk-aware behaviors and decision-making with performance. As such, organizations need a rigorous and controlled process to document the algorithm's purpose as well as needs and goals for the organization. This should be included in an organization's AI architecture document and related software development processes.

Along with clear visibility for top executives and board members, governance of underlying data is key to effective ERM framework. For successful implementation, organizations must evaluate what data is needed to develop AI. AI algorithms use data to train and create a novel model. The models predict future outcomes as they receive new data. Necessary data governance considerations, drawing from core values, may include 1) representation of the appropriate population for the AI use case and reduction of bias; 2) clear rules for using and disseminating data, including privacy in data collection as well as disclosure of use and disposal; and 3) ways to secure data assets.

AI and the models that make it work also have to be closely monitored across an organization. In designing and implementing AI, six key dimensions may help safeguard ethics and build a trustworthy AI strategy for the company that people can embrace. Although currently there is no authoritative framework for AI ethics, Deloitte's Trustworthy AI™ Framework can serve as a means to understand and assess risks and ethical considerations that are specific to AI and can be a valuable lens to complement the COSO ERM Framework, especially as it relates to governance and performance. Organizations can use it to help determine and monitor ongoing risks.

Figure 6. **Deloitte's Trustworthy AI™ Framework**



Deloitte's Trustworthy AI™ Framework (see Figure 6) includes the following:

- **Fair and impartial** – Assess whether AI systems include internal and external checks to help enable equitable application across all participants.
- **Transparent and explainable** – Help participants understand how their data can be used and how AI systems make decisions. Algorithms, attributes, and correlations are open to inspection.
- **Responsible and accountable** – Put an organizational structure and policies in place that can help clearly determine who is responsible for the output of AI system decisions.
- **Robust and reliable** – Confirm that AI systems have the ability to learn from humans and other systems in order to produce consistent and reliable outcomes.
- **Privacy** – Respect data privacy and avoid using AI to leverage customer data beyond its intended and stated use. Allow customers to opt in or opt out of sharing their data.
- **Safe and secure** – Protect AI systems from potential risks (including cyber risks) that may cause physical and digital harm.

Points to Ponder

- Does the organization have an integrated AI governance program?
- How are ethical considerations factored into AI implementation? Should there be a chief ethics officer to govern ongoing monitoring of AI?
- Does the organization have a chief risk officer, data officer, or equivalent risk leader to help with risks associated with enterprise-wide AI initiatives?
- Does the board have a member who is a technology or AI expert?
- What board-level approvals or consultations happen around AI implementation and changes post-implementation?

STRATEGY & OBJECTIVE-SETTING

Every organization has a strategy for bringing its mission and vision to fruition, and to drive value. Organizations should integrate ERM with strategy setting to gain insight into the risk profile associated with the organization's strategy and business objectives. The Strategy & Objective Setting component and following principles of the COSO ERM Framework serve as the basis for this section of the paper:

6. **Analyzes business context**
7. **Defines risk appetite**
8. **Evaluates alternative strategies**
9. **Formulates business objectives**

Organizations should establish a strategy and business objectives in relation to AI. According to Deloitte's 2020 Chief Strategy Officer Survey, 51% of respondents indicated AI was important for their organization's strategy, but 17% felt that their organization had capabilities to execute AI-related strategies.[12] With an understanding of business and strategic context, an organization's leaders can understand internal and external factors that impact risks for their AI initiatives. Important components are classification of current or potential uses of AI and related data, and evaluation of potential exposure to AI use.

Use of AI models that do not align with an organization's values can hurt strategic objectives. There are numerous examples where organizations have used AI models that produced outputs that resulted in different or unequal treatment of participants based on protected characteristics (e.g., gender, race, etc.). These incidents suggest an inadequate focus on identifying and addressing issues related to fairness and transparency during both AI development and ongoing monitoring.

Defining risk appetite enables organizations to align risk identification, assessment, and response to business strategy. For a more in-depth discussion on defining risk appetite, refer to COSO's "*Risk Appetite – Critical to Success: Using Risk Appetite to Thrive in a Changing World*".[13] An additional consideration in developing risk appetite is benchmarking against industry peers. An organization's risk appetite is also a key consideration in facilitating risk-informed decision-making as it relates to AI. There's no way to eliminate risk entirely, thus the organization must determine its risk appetite and assess how much risk is acceptable when identifying and evaluating investments in AI. There's a risk/reward trade-off to consider. Organizations adopting AI are making significant investments in development and implementation and must align their AI risk management with broader risk management efforts. According to Deloitte's latest State of AI in the Enterprise report, 43% of surveyed seasoned AI adopters are pursuing such alignment.[14]

Achieving Outcomes with Lower Risk

AI can create significant efficiency and benefits for an organization. Organizations are using AI to monitor components of a variety of manufacturing processes. For example, a manufacturer may use AI to predict when conveyor belts are likely to fail. Instead of using current fail-data for belts, the manufacturer may use AI via temperature measurements, video cam feeds, and other novel variables to identify fail points, thus creating new data to build a useful model. The training data and the data within the model itself may help drive efficiency for the manufacturer's processes. This AI example illustrates a relatively common AI use case with reduced risk.

[12] Deloitte, 2020 Chief Strategy Officer Survey, a Monitor Deloitte and Kellogg School of Management study.

[13] Committee of Sponsoring Organizations of the Treadway Commission (COSO), "Risk Appetite – Critical to Success: Using Risk Appetite to Thrive in a Changing World," May 2020. https://www.coso.org/Documents/COSO-Guidance-Risk-Appetite-Critical-to-Success.pdf.

[14] Ibid., Figure 2 on page 7.

The interplay between developing strategy and risk appetite is a key input to an organization's risk assessment. Informed by its risk assessment, an organization determines its responses to identified risks. An organization's response should include setting up control activities (e.g., inventorying, benchmarking, and trends analysis) that manage the identified risks. Post-implementation, it's important to measure outcomes to determine whether business objectives have been achieved with lower risk. Only about 34% of surveyed AI adopters are maintaining a formal inventory of all AI implementations.[15] Without maintaining such an inventory, it's difficult to monitor and evaluate potential exposure from AI use cases.

Drawing from a risk tolerance definition, which is one of the key parts of the COSO ERM Framework, helps to establish key performance and risk indicators around AI to monitor performance of algorithms over time. Setting up key performance and risk indicators and tolerance levels while the algorithm is being developed helps create a performance baseline by which to articulate trust. Reporting of such metrics brings transparency among stakeholders, which may help improve the performance of algorithm and integrity of the underlying input data.

Points to Ponder

- Does the organization use strategic risk assessment techniques like scenario planning and assumptions testing for AI programs?
- Are AI capabilities used for identifying emerging risks and seeking stakeholder feedback about products, services, and brand?
- Do AI initiatives support risk analytics to monitor risks?
- Do AI risk assessments consider the risks *and* rewards associated with each AI use case and factor these trade-offs into both go/no-go decisions as well as design and purpose of relevant AI models?

[15] Ibid., based on average from Figure 9 on page 15.

Identifying, assessing, and responding to risk are key activities that organizations should undertake to support the achievement of the organization's strategy and business objectives. Risk, especially AI-related risk, emanates from a variety of sources, and organizations need to adopt a range of responses across the organization and at all levels. The Performance component and following principles of the COSO ERM Framework serve as the basis for this section of the paper:

10. **Identifies risk**
11. **Assesses severity of risk**
12. **Prioritizes risks**
13. **Implements risk responses**
14. **Develops portfolio view**

Organizations should not implement AI applications without addressing their trustworthiness. To unlock the full potential value, AI models should be built with trustworthy AI in mind and include performance considerations that help to make AI robust, reliable, safe, and secure while maintaining privacy.

Not all AI models have the same risk profile. Organizations will need to perform risk assessments to solidify each business case. The identification of risks related to AI initiatives is also necessary to evaluate exposure and identify opportunities for a higher adoption of AI for value creation. Organizations also need to prioritize risks by assessing AI models and determining the level of accuracy, reliability, and transparency required for the related use case(s). AI models that require a high level of accuracy, reliability, or transparency to achieve success likely have a higher risk profile. In addition, an AI model that is being used to provide a suggestion for a low-impact decision (e.g., which song to play next) will have a lower risk profile than an AI model that is being used to automate decisions previously made by humans (e.g., deciding on underwriting terms for an insurance policy).

Organizations should consider the severity and priority of the risk as well as the business context, business objectives, and performance targets of the AI model in selecting and deploying a risk response. Risk responses related to AI models generally fall within the following categories:

- **Accept:** No action is taken to change the severity of the risk. This response is appropriate when the risk to strategy and business objectives is already within risk appetite. Risk that is outside the organization's risk appetite and that management seeks to accept will generally require approval from the board or other oversight bodies.
- **Avoid:** Action is taken to remove the risk, which may mean not using the AI model, limiting the scope of use of the AI model, or modifying the functionality of the AI model to limit complexity.
- **Pursue:** Action is taken that accepts increased risk to achieve improved performance. This may involve expanding the scope of use of AI models or modifying the functionality of the AI model to increase complexity. When choosing to pursue risk, management understands the nature and extent of any changes required to achieve desired performance while not exceeding the boundaries of acceptable risk tolerance.
- **Reduce:** Action is taken to reduce the severity of the risk. This involves establishing business processes and controls that reduce residual risk to an acceptable level aligned with the organization's risk profile and appetite. (Actions organizations may take to reduce risk associated with AI models are described below.)
- **Share**: Action is taken to reduce the severity of the risk by transferring or otherwise sharing a portion of the risk. A common example is outsourcing development, implementation, or monitoring of AI models to specialist service providers.

Although it's not possible to completely avoid AI risk, there are actions organizations can take to reduce risk. One is to develop a testing regime for developed or implemented AI solutions and apply the testing regime throughout the AI solutions' lifecycle. Approximately 40% of surveyed AI adopters currently conduct internal audits and testing of their AI implementations.[16]

Artificial Intelligence Sometimes has Unintended Consequences

The performance of algorithms must be comprehensively assessed for fairness, transparency, and robustness. They have the potential to drift from the original strategic intention as they ingest more data.

- **Fair and impartial:** Is there bias toward certain groups, justified differential treatment of groups, or a fair representation of relevant populations?
- **Transparent and explainable:** What are the main contributors that influence model output and how does each input factor influence the result?
- **Robust and reliable:** Will the model remain stable in the future and generalize well to unseen data or is there a risk of future bias as the model receives new data?

Key actions in assessing performance of AI models include but are not limited to:

- **Risk review** helps identify risk factors, including cybersecurity, data risks, bias, and ethics, that could prevent or sub-optimize the goals of successful AI implementation. A portfolio view of risks associated with all AI projects should be reviewed with senior management and the board of directors. A key aspect of this review is implementation of risk responses where each response and the residual level of risk should be carefully evaluated against the risk appetite definition.
- **Data review** helps evaluate quality and integrity of data and its impact on AI models and their outcomes. Data review also helps identify correlations between variables. For example, does age and/or body mass index correlate to getting cancer? Organizations can perform multivariate analysis of underlying data to identify historical sources of bias that may be used as input to the algorithms.
- **Model review** tests outcomes using the following actions:
  1. Analysis of the algorithm's functional form and parameters to understand possible problems in the decision-making process.
  2. Assessment of algorithm performance on real data to test for hidden biases resulting from complex correlations or other unexpected sources of real-world error. Correlation is important because it helps identify the presence of an association between a protected variable (e.g., gender, race, etc.) and variables that may serve as potential proxies for a protected variable used in the model. If such a relationship exists, the model may contain bias. Statistical significance indicates that the relationship between these variables is not caused by random chance.
- **Implementation review** helps ensure an AI algorithm is working correctly. This review helps assess whether the algorithm will continue to be robust, effective, and fair in the future, and identifies potential risks.
- **Post-deployment review** looks at algorithms on a repeated basis. It's necessary to periodically assess model performance and fairness after deployment. This assessment likely requires a monitoring mechanism that continuously tests the underlying data and functionality of the model.

Complications When AI models Perform Outside their Test Environment

Testing the performance and outputs of an AI model includes considering unexpected data/behavior or changes within the data to evaluate the reliability of the outcomes from the AI model. Depending on how the AI model is designed, the introduction of unexpected data/behavior or changes in the data may result in the AI model producing incorrect/harmful outputs or not functioning at all. There can be significant consequences for organizations that implement AI models that are not robust and reliable.

[16] Ibid., based on average from Figure 9 on page 15.

Keep in mind that AI programs can be hacked like any other data source or company. Deloitte found that 62% of respondents have significant concerns about cybersecurity vulnerabilities but only 39% are addressing these risks.[17] To keep AI applications and related data safe and secure — a trustworthy AI pillar — organizations must implement and maintain a model-version control methodology, including maintaining a baseline version of the AI model and tracking each subsequent version and the changes made to it to enable auditability, transparency, and reproducibility of the AI model. The data-version control methodology provides the foundation. Organizations should establish incremental preventative, detective, and monitoring controls around the model as well as data used to train the underlying algorithms within the model to prevent and detect unauthorized or malicious changes. Due to the computing power necessary to drive many of these models, the processing takes place in the cloud, which introduces third-party reliability and privacy concerns as well.

Furthermore, policies are required that address securely retaining personal data (encryption, anonymization, etc.), data disposal and communicating what is obtained, how it is used, and how it is maintained. Deloitte found that 57% of respondents have significant concerns about the consequences of using personal data without consent, but only 37% are addressing these risks.[18] Privacy is an important pillar for achieving trustworthy AI.

Rules about when further review is necessary must be established. Organizations should define deficiencies, performance measures, and thresholds that require further investigation or escalated review. In addition to Performance, these rules support the Governance & Culture and Strategy & Objective Setting components of the COSO ERM Framework. Key inputs include but are not limited to the following items:

- The organization's definition of success (not just financial or operational) for AI initiatives and related AI models
- Identified risks to achieving that success
- Controls designed and implemented to manage those risks

As part of responsibility and accountability, one of the pillars of trustworthy AI, organizations need to define and execute processes to monitor for continued success. They also should define and execute remediation when success is not achieved. People need to be specifically responsible for those activities. To help those responsible, the architecture needed to support monitoring, and escalation can be built into the AI platform. Automation can help facilitate the monitoring and escalate reviews to designated people in real time.

Points to Ponder

- **Do AI model performance reviews include assessing and managing risks to improve results?**
- **Are key risk and performance indicators for AI applications monitored through executive dashboards and reported to authorized data users?**
- **How much confidence is there that the AI application and related controls are operating as intended and generating the right information for decision-making?**

[17] Ibid., Figure 8, page 14.

[18] Ibid., Figure 8, page 14.

REVIEW & REVISION

In an ever-changing business environment, an organization's strategy or business objectives and ERM practices and capabilities may change over time. Specific to the realm of AI, the ongoing changes in capabilities and expansion of uses require an organization to continually assess its ERM practices and capabilities, and revise them if necessary. The Review & Revision component and following principles of the COSO ERM Framework serve as the basis for this section of the paper:

15 **Assesses substantial change**
16 **Reviews risk and performance**
17 **Pursues improvement in ERM**

As mentioned previously in this paper, organizations are increasingly adopting AI and are anticipating organizational and industry-wide transformation from their investments in AI. In addition, regulatory agencies and governments have enacted and are deliberating over additional regulations pertaining to the use of AI and related data. These developments may lead to substantial changes, including functionality of AI models, which may result in new or changed risks. Such development may also affect ERM as well as the achievement of strategy and business objectives. An iterative process that can affect several components of ERM involves identifying substantial changes and their effects, and responding to those changes.

Reviewing ERM practices and capabilities along with the organization's performance relative to its targets helps enable organizations to monitor how their AI applications increase value and will continue to drive it. Management needs to test and monitor AI and machine learning applications to help ensure the applications work as they're intended. Ongoing monitoring of performance and risks helps assess if AI is delivering on its intended objectives and establishes a cycle of risk-informed decision-making.

A risk taxonomy focused on the AI model and related initiative should be developed to address the universe of AI risks. Risk management teams must help develop the taxonomy that will guide risk identification and assessment efforts. Organizations can use the COSO ERM Framework and other guidance to help identify, assess, prioritize, and monitor AI-related risks. Assessing the AI model's achievement of objectives demonstrates the value of risk management and highlights opportunities for improvement.

Key performance and risk indicators are important to maintain for the long term because algorithms change as they learn and may produce unintended consequences in the future. Furthermore, even the best-intentioned algorithms are subject to bias or issues related to reliability. Simply omitting personally identifiable information (PII), such as race and gender, may not be sufficient. Continual monitoring and testing of algorithms is necessary especially as data used by algorithms and trends within the data change over time.

The three lines of defense model can be used whereby each stakeholder can play a role in review and revision of AI applications and their performance. The first line, guided by ERM, can proactively identify and address risk factors for AI, while ERM (the second line) can collaborate with the first line and make risk assessments effective, dynamic, and actionable. ERM can also collaborate with the first line stakeholders to present insightful risk reports and recommendations to the leadership. Internal audit, using risk-based approach, can play an independent reviewer role and critically assess AI applications for business performance and risk management goals.

Serious Issues Can Arise When Performance Isn't Reviewed and Monitored

For example, AI models are increasingly being used within healthcare to assist in diagnosing conditions and providing medical advice. If organizations or medical professionals do not properly monitor the performance of these models, they may not identify and correct cases where the AI models provide inaccurate diagnosis or medical advice. Failure to identify and correct inaccurate results may lead to medical harm, patient concerns, and questions about the process for building the related AI models.

Points to Ponder

- Does the organization perform a portfolio review of all AI programs to understand synergies and risks at an aggregate level?
- Does a chief risk officer participate in AI performance reviews to share risk management perspectives?
- Are findings, both positive and negative, shared with the members of senior management and board of directors in such reviews?
- Does senior management take appropriate remedial actions to address any negative findings?
- Do you have a multidisciplinary risk management team that can help with AI model risk mitigation planning?

INFORMATION, COMMUNICATION, & REPORTING

Organizations are continually challenged to use the enormous quantity of data generated coupled with the increasing concerns over privacy and security of data and transparency of related AI models. In this environment, it is important that organizations provide the right information, in the right form, at the right level of detail, to the right people, in a timely manner. The Information, Communication & Reporting component and following principles of the COSO ERM Framework serve as the basis for this section of the paper:

18 **Leverages information and technology**

19 **Communicates risk information**

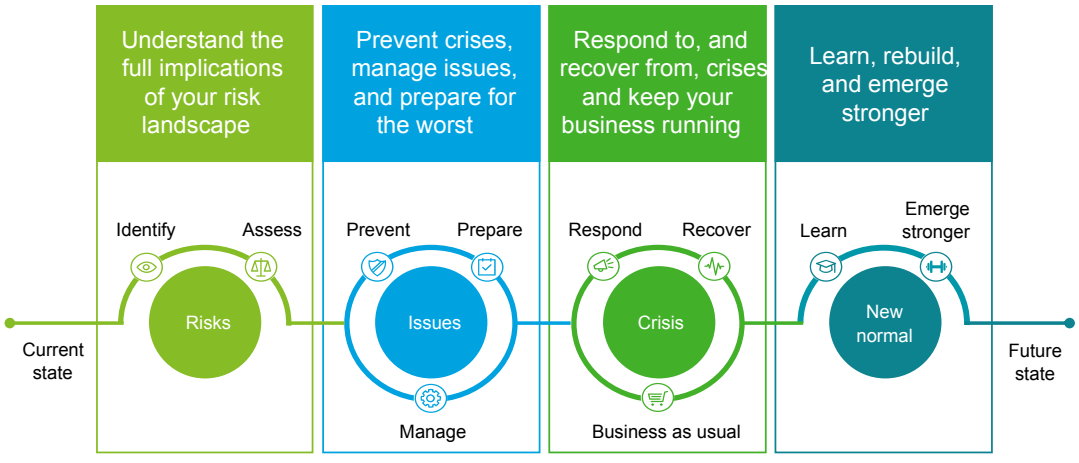
20 **Reports on risk, culture, and performance**

Reports on risk, culture, and performance use IT systems to capture, process, and manage data and information. Management uses that information to inform and support risk management, including risk management related to AI models. A reporting process is needed to inform internal and external stakeholders about the performance, benefits, and potential risks of AI models. The reporting process also considers how, when, and how often stakeholders will receive the information. In building an organization's resilience, an understanding of the risk landscape is needed, and a unified AI risk report should be compiled for executive management and board members to aid their oversight efforts. This report may include updates regarding key performance measures and risk indicators for performance of the organization's AI models, as well as results from key oversight and monitoring processes. Timely communication of results, including unexpected findings, is vital for identification and resolution of issues before they grow into larger problems.

To prevent crises, manage issues, and prepare for worst-case scenarios that may emerge from undesired performance or incidents related to AI initiatives, a crisis communications response framework and protocols should act as a guide. (See Figure 7) Such a crisis communications playbook will spell out how an organization should respond to control the impact and exposure from any incidents while keeping the business running. It should also include steps to assist recovery.

Data around stakeholder reactions is an important component of rebuilding and emerging stronger following a crisis. These responses will help inform AI strategy and implementation and assist the organization in meeting expectations for transparency.

Figure 7. **Building Resilience**



AI Use in the Spotlight

AI is increasingly becoming a large part of organization's business operations. In recognition of investors' increasing interest in AI use, several large technology-based companies have included disclosures in their 10-K filings that outline how AI models currently impact business operations and their potential impact in the future.

Points to Ponder

- Is there a crisis response plan in place?
- What AI program performance reporting is disseminated to stakeholders and to the public?
- Do executives and oversight bodies within the organization receive relevant performance information around AI programs?

SUMMARY REMARKS

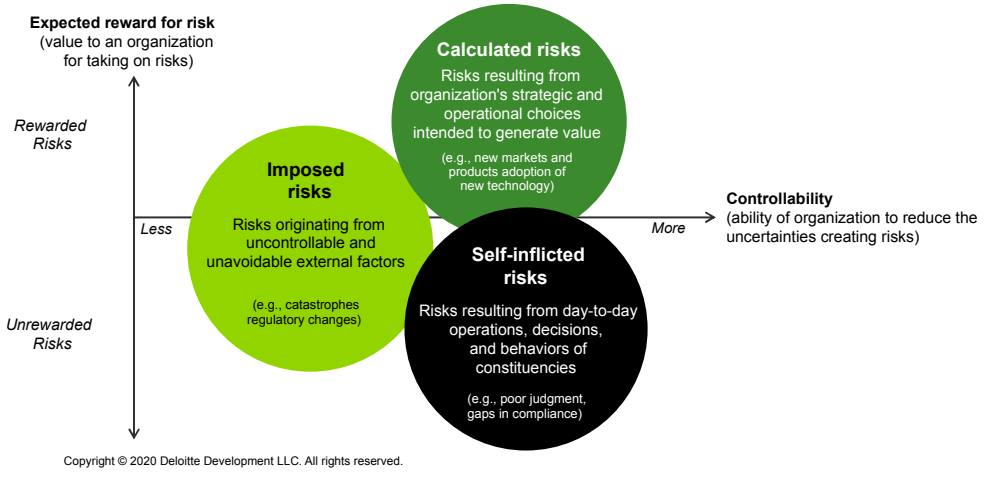
To realize AI's value and take advantage of its potential, organizations must align risk management with their strategy and execution of their AI initiatives. The COSO ERM Framework can help organizations develop integrated governance over AI, manage risks, and drive performance to achieve strategic goals. By implementing integrated governance over AI, organizations can have better information about relevant risks. This may support an increase in the range of opportunities and flexibility to take calculated strategic risks and become nimbler and more adaptive in planning and executing their AI initiatives. Although not authoritative, the Deloitte Trustworthy AI™ Framework can help organizations think through the risks when implementing COSO's ERM Framework for AI.

Figure 8. **COSO *Enterprise Risk Management – Integrating with Strategy and Performance Framework***

Through ERM, informed by the COSO ERM Framework, organizations can reduce performance variability and improve the likelihood of success for their AI initiatives. By identifying signals to correct course early, organizations can increase positive outcomes, reduce negative surprises, and improve resilience to risk. Risk-informed resource allocation can also be improved and, by understanding its risk, the organization may be better equipped to deliver return on investment and meet stakeholder expectations. Furthermore, by implementing ERM, organizations can refine and adapt their innovation initiatives to support their strategies in a rapidly changing business environment.

Properly implemented risk management can help organizations take advantage of calculated risks with high rewards, manage inherent risks and help significantly decrease self-inflicted risks. (See Figure 9).

Figure 9. **An ERM Program Helps Organizations Achieve Success Related to Their AI Initiatives**



AI solutions need to be trusted, tried, and true. Trusted – because ERM is transparent by nature and it helps keep an organization abreast of its risks and opportunities. Tried – in that models are continually tested and vetted to verify they are operating as intended. And True – governance, risk management, testing, and monitoring regimes help models to operate in ways that reflect the organization's values and protect its reputation. The COSO ERM Framework, when considered appropriately, can result in trusted, tried, and true AI.

Call to action: Five next steps to consider based on the COSO ERM Framework

Use the COSO framework and underlying components and principles to establish a trustworthy AI program. Here's how to get started:

1. **Establish governance structure for AI program.** Determine when and how the organization will use AI and define the purpose and objectives of proposed AI initiatives. This includes evaluating applicable ethical considerations. Bring various AI initiatives across your organization under an overall AI program and a governance structure providing visibility to senior management and board of directors. Identify a senior executive to lead your AI program and provide risk and performance oversight.

2. **Get an AI risk strategy together.** Collaborate with stakeholders to draft an organization-wide strategy to manage the strategic, technical, regulatory, and operational risks of AI. Ensure that your organization has the AI technical experience to execute the AI risk strategy. The strategy should define roles, responsibilities, controls, and mitigation procedures.

3. **Take the initiative with AI risk assessment.** For each AI model your organization uses, gauge the potential impact of suboptimal strategic outcomes, operational failures, or bias. Also, evaluate how the algorithm manages and uses data and whether it introduces any unintended bias. For business processes that integrate with AI, look for vulnerabilities and see how likely they are to occur, then record known risks and corresponding controls.

4. **Develop a portfolio view of risks and opportunities for AI initiatives.** Chief Risk Officer and AI leader can work together to proactively review AI models for risks pertaining to bias, tampering, and model malfunction. They should report a portfolio view of AI risks to senior executives and board of directors for awareness and decision-making support.

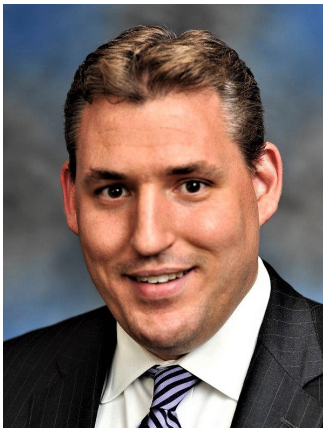
5. **Lay out an approach to manage AI risks and report to stakeholders for transparency.** This includes evaluating risk-reward trade-offs for AI initiatives and resource allocation. Consider assembling a team of AI model risk experts to offer leading practices, objectivity, and risk response methodologies. Establish key performance and risk metrics to measure goals such as efficacy, fairness, and transparency of each model. For each metric, set thresholds that would trigger off-cycle model reviews and corrective actions. Develop reporting dashboards for executives and boards of directors, as well as disclose AI performance and risk management actions to external stakeholders for awareness.

ABOUT THE AUTHORS

Keri Calagna, Risk & Financial Advisory Principle, Deloitte & Touche LLP

Keri is a leader in Deloitte's Cyber and Strategic Risk practice. With more than 25 years of risk experience, Keri helps organizations evolve their culture, capabilities, and processes to create integrated risk programs that help grow the business, accelerate performance, improve resilience, and achieve strategic goals. Throughout her career, she has helped businesses evaluate, manage, and monitor a wide spectrum of risks including financial, operational, reputational, regulatory, enterprise, strategic, and technological risk. Keri commonly advises boards and executive leadership teams on the design and roll-out of enterprise-level risk governance, monitoring, and reporting and helps executives align and mobilize around the top-most risks to their organization.

Keri received a BS and MBA in Entrepreneurship from Rensselaer Polytechnic Institute's Lally School of Management & Technology.



Brian Cassidy, Audit & Assurance Partner, Deloitte & Touche LLP

Brian is the US Audit & Assurance Artificial Intelligence/Algorithms leader with diverse experience providing audit and advisory services to Fortune 500 companies. A leader who brings strong technical, risk management, communication, and organizational skills, he focuses on providing audit, accounting, and advisory services to public and private companies in the financial services sector. Brian's experience crosses a wide range of industries in the financial services sector, including banking (brokers/dealers), investment companies, business development companies, and alternative funds, including private equity, hedge, and real estate. He also leads Deloitte's efforts in the Algo/AI assurance area as emerging technologies continue to impact clients and the marketplace.

Brian is a member of the American Institute of Certified Public Accountants (AICPA), The Pennsylvania Institute of Certified Public Accountants (PICPA), and The New York State Society of CPAs (NYSSCPA). Brian received a BS in Accountancy and BS in Business Administration from Villanova University.

Amy Park, Audit & Assurance Partner, Deloitte & Touche LLP

Amy is the Ideation leader for Deloitte's Accounting Advisory & Transformation Services practice. In this role, Amy leads the development of ideas into potential new service offerings that can enhance the value Deloitte can bring to the marketplace, including areas of emerging technologies and expanded assurance, such as AI, algorithms, blockchain and digital assets. She is also a partner in Deloitte's National Office Accounting and Reporting Services and specializes in technical accounting matters in consolidation, financial instruments, and accounting for digital asset transactions.

Amy is a member of the American Institute of Certified Public Accountants and serves on the AICPA's Digital Assets Task Force, focusing on accounting matters related to digital assets. She has more than 17 years of experience in public accounting, including a practice fellowship at the Financial Accounting Standards Board, and has served public and private companies in the banking and securities and digital assets industries.

ABOUT COSO

Originally formed in 1985, COSO is a joint initiative of five private sector organizations and is dedicated to helping organizations improve performance by developing thought leadership that enhances internal control, risk management, governance, and fraud deterrence. COSO's supporting organizations are the American Accounting Association (AAA), the American Institute of Certified Public Accountants (AICPA), Financial Executives International (FEI), the Institute of Management Accountants (IMA), and The Institute of Internal Auditors (IIA).

ABOUT DELOITTE

COSO

Committee of Sponsoring Organizations
of the Treadway Commission

coso.org

REALIZE THE FULL POTENTIAL OF ARTIFICIAL INTELLIGENCE

APPLYING THE COSO FRAMEWORK AND PRINCIPLES TO HELP IMPLEMENT AND SCALE AI

COSO

Committee of Sponsoring Organizations of the Treadway Commission

coso.org