ECI AI WORKING GROUP: NAVIGATING ARTIFICIAL INTELLIGENCE TO OPTIMIZE RISK MANAGEMENT AND MINIMIZE ETHICAL IMPLICATIONS

ECI

This report is published by the Ethics & Compliance Initiative (ECI).

Ethics & Compliance Initiative®

ISBN: 979-8-3507-3376-1

ECI
8409 Lee Highway #2175
Merrifield, VA 22116-8239

Tel: 703.647.2185 | FAX: 703.647.2180
www.ethics.org | **research@ethics.org**

ABOUT THE ETHICS & COMPLIANCE INITIATIVE

The Ethics & Compliance Initiative (ECI) empowers organizations to build and sustain High Quality Ethics & Compliance Programs (HQP®). ECI provides leading ethics and compliance research and best practices, networking opportunities and certification to its membership.

Acknowledgements:

We are grateful to the following members of our Working Group for their many hours of effort in compiling this report:

CHAIR:

Cindy Cetani
Indivior

MEMBERS:

Tiffany Butler
World Concern

Liam Dagless
GSK

Chris Dorsey
National Grid

Colleen Dorsey
University St. Thomas

Diana Floyd
Alta Resources

Julie Gershman
Prudential Financial

Jamilla Jean
Baker McKenzie

Kim Kuennen
CUNA Mutual Group

Blair Marks
Lockheed Martin

Kim Miller
Rolls-Royce

Page Motes
Dell, Inc.

Mary Ann Nord
Leidos

Margaret Huggett
BP

Dinesh Relwani
The Boeing Company

Paul Zikmund
Baker Tilly

ECI AI WORKING GROUP: NAVIGATING ARTIFICIAL INTELLIGENCE TO OPTIMIZE RISK MANAGEMENT AND MINIMIZE ETHICAL IMPLICATIONS

Table of Contents

Introduction

AI, NLP, ANN, etc.... As a compliance professional with a broad scope, limited free time, and a need to prioritize your focus, do you really need to add more acronyms to your vocabulary? Aren't artificial intelligence and the technology surrounding it the responsibility of our information technology colleagues? The answer rests in the value proposition.

Being knowledgeable about artificial intelligence is a valuable addition to your toolbox and will:

- **enable you to optimize your resources by leveraging data for targeting risks and strategic decision-making.**
- **position you as a valued strategic business partner as you provide support and counsel to your colleagues across the company who embark on artificial intelligence solutions to drive business and solve challenges in their areas of responsibility.**
- **enhance your ability to uphold your company's standards of ethics as artificial intelligence solutions are explored.**
- **help you further evolve in line with increasing regulatory authority expectations to mine vast information sources for effective risk management.**
- **require that you fully understand how it all works, including both the upside and downside (or ethical implications).**

This paper will help you navigate this topic and covers the following areas:

1. Introduction to artificial intelligence
2. Leveraging data
3. Practical applications
4. Ethical implications

Key takeaways include discussions of the essential interaction between humans and machines; the need to take a thoughtful, phased approach to AI; and the planning and resources required to produce reliable results.

Introduction to Artificial Intelligence

Most people have heard of artificial intelligence (AI) or machine learning, but unless you're an engineer or a computer programmer, it's hard to get your head around what exactly it means—let alone how to design it, monitor it, and control it so it is used appropriately. Yet the importance of understanding it today and in the future cannot be overstated. Let's navigate some of this nomenclature and how it all works.

The journey to understand the true meaning of AI might best begin by understanding how it got to be such a big deal in the world we live in today. After all, it has been around since the 1950s—so why the hyper-focus now? There are three main reasons:

1. There are **massive amounts of data now available**. Think about all the sources of data available today: health data; employment data; personal data like where we live and work, the house of worship we attend, what kind of food we eat, the route we take to work, etc. The uses for this kind of data are as endless as the data itself—for example, the data can be used to determine whether you are at risk for lung cancer, whether you are a good candidate for a loan based on your financial status, whether you are susceptible to insider trading and whether there are accidents on the street in which you live.
2. There are **unprecedented amounts of available computing power**. What was once stored on servers at great expense can now be stored in the cloud for a fraction of the cost.
3. We have **access to sophisticated and breakthrough algorithms, tools and frameworks**.

The combination of these factors has set into motion a heightened level of attention by companies of all sizes and across all industries. Companies want to learn how to create or make use of technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways similar to how people do—but faster. In other words, they want to leverage AI.

These desired capabilities are generally associated with what AI can do. Let's look at each of them in the context of an example:

1. **LEARNING**. Traditionally, the learning capability associated with AI began with programmers creating algorithms (see below). This was a good way to gauge what computers were capable of learning. A developer would create a program, deploy it, update the algorithm to build in enhancements and redeploy the program. AI enables the machine to "learn" over time without this direct human intervention. For example, at Infervision in China, there were not enough radiologists to review the 1.4 billion CT scans produced annually to detect early signs of lung cancer.[1] They faced a time-consuming, laborious and, frankly, tedious task of reviewing hundreds of scans daily. There was a risk of human error, often caused by fatigue and resulting in mistakes, misdiagnoses and important diagnoses being missed completely.

> "Computers are also capable of perception. This occurs when a computer recognizes images and applies an interpretation across the available data."

To address this situation, Infervision applied a deep-learning solution leveraging algorithms to augment the work of the radiologists and allow them to diagnose cancer more accurately and efficiently.

2. **PERCEPTION**. Computers are also capable of perception. This occurs when a computer recognizes images and applies an interpretation across the available data. In our China example, the computer was provided real-life data from previous scan images of confirmed lung cancer to increase its accuracy in identifying signs of cancerous growth in lung scans. Perception in this context is the ability of the computer to accurately interpret data.

3. **COGNITION**. Cognition occurs when the computer applies "reason" over data to identify the best possible action to take given a certain output. Again, using our lung cancer AI product as an example, this might entail the program providing recommended actions for the radiologists to consider given the information the system is interpreting.

Another aspect of understanding AI relates to how it is created. In this vein, it might be useful to describe what it means to write an algorithm and how that ties into machine learning and deep learning.

An algorithm is a set of well-defined instructions or commands written clearly in sequence to solve a particular problem. It can then be used in different programming languages (think JavaScript, Python, Ruby, etc.). Here's an example of an algorithm to add two numbers entered by the user:

Step 1: Start

Step 2: Declare variables num1, num2 and sum

Step 3: Read values num1 and num2

Step 4: Add num1 and num2 and assign the result to sum

sum ← num1+num2

Step 5: Display sum

Step 6: Stop

Simple algorithms like this example are useful and have their place, but they are not capable of solving more complicated problems such as facial recognition.

Machine learning, often confused with AI, takes a different approach. Machine learning looks at historic data to extract patterns from prior situations or trends. Instead of the programmer manually creating the algorithm by writing commands, the computer uses **training data** to create the algorithm. Imagine that we want to develop a program that recognizes a dog (the output) across many images of both dogs and cats. The training data "inputs" would be thousands of pictures of dogs. Once we have the results of that training, we have a **model**. We can then put new varied data into the model to get the desired output (in this case, the computer's trained guess at identifying images of dogs). This is similar to our lung cancer AI model above. Machine learning reads mined data to create an algorithm, adapts when exposed to new data and creates new algorithms through AI.[2]

> "Machine learning reads mined data to create an algorithm, adapts when exposed to new data, and creates new algorithms through AI."

Natural Language Processing (or NLP) is another form that combines language, machine learning, and artificial intelligence. This has become a more common form, and we participate in it via Siri on our iPhones or other home devices where we use our spoken language (vs. typing) to interact with devices.

To further understand how advanced AI has become, it is useful to have a basic understanding of linear algorithms. A linear search is the most basic of searching algorithms. It essentially looks down a list one item at a time. Think about searching through a paper phone book, one name at a time, until you find who you are looking for. Linear searches have been used for decades. Developers have now developed what are called **artificial neural networks (ANN)**. A key to understanding ANN, at least at a basic level, is to understand how our brains process information. The brain can take multiple inputs and further process them to infer hidden as well as complex non-linear relationships. It is also capable of learning from its past mistakes. ANN attempts to duplicate how this brain neuron system functions.

With a general understanding of what AI is and how it attempts to mimic the human brain, it is hopefully a bit easier to see how far-reaching it can be in our everyday lives. You can see from these examples that some form of human intervention is involved on both the front and back ends. This is essential for the accuracy of what the model produces and to avoid unintended bias from being built into the model. Keep reading as we will cover this important topic later in the paper.

Leveraging Data

How can you structure the data to get desired and reliable insights later? Start with a data audit.

You now know that machine learning is a form of pattern recognition that requires large amounts of data to function. What data a company uses will be determined by what outcomes the company wants from that data. Historic data can be used to effectively predict future outcomes when the right data is first identified and grouped. This requires the creation of data sets to find solutions to identified problems. In other words, you will need to identify sets of data with common "signals" about information important to you.

> "Where do you start? Companies must first identify and prepare their available data. An individual or a team will need to determine which data is useful to them."

Where do you start? Companies must first identify and prepare their available data. An individual or a team will need to determine which data is useful to them. The first step to get your organization AI-ready is to perform a **data audit**. Available data might include information as straightforward as the number of employees or merchandise available in a store or as complex as customer demographics. The diversity of data is infinite, which is why it is important to prioritize based on strategic goals.

The extra effort, and challenge, at this initial stage will be to find data that might be "siloed." Every company has data in Excel spreadsheets, on personal desktops, across departments and across business units. Further, this data might include "unstructured data"—data that is embedded in text-heavy documents and materials—as well as "structured data" like financial numbers. Both unstructured and structured data may be filled with dates, personally identifiable information and other information that needs to be identified for appropriate handling. All the siloed data must then be gathered into a common repository.

Key steps during a data audit include:

1. **DATA LAKE**: This is a raw storage repository of data. Unlike data that is organized in files or folders, a data lake uses "flat architecture." This means that data scientists, developers and other data specialists can access raw data with their desired set of analytics later. A key takeaway here is the need to secure the noted data specialists either internally or externally.

2 **DATA INVENTORY**: A manual input of data to create an inventory could be as simple as inserting data into a spreadsheet where you record (1) repository location, type, owner and accessibility of data, (2) the content type and its identifier, and (3) any personal information, aligned with the EU's General Data Protection Regulations on processing personal data or with other relevant personal data regulations as necessary. Company information that is not already in the form of data will need to be "datafied." For example, social media sites datafy friendships and posts, locations, and times of posting. Algorithms will be created from the datafied information. These algorithms are simply shortcuts used to tell computers what to do—for example, how data will be tagged and inventoried.

This tagging or taxonomy of data will help to ensure that disparate sources all call the same data by the same name. Streamlining a tagging system for data at this stage will help guide the various other departments and functions to tag data appropriately and in the same manner going forward. For example, imagine you have to sort through a million files for the word "blue." Even if it only took you one second per file, you'd have to sort for more than 11 days straight without stopping to sleep or eat. However, if you taught a computer to recognize the word "blue" using an algorithm, it could do the work for you—and, given enough processing power and proper algorithmic tuning, it could probably accomplish the task in a few seconds.[3]

3 **DATA MAP**: The data map complements your data inventory. Simply stated, the data map sorts data into a searchable and easily identifiable format. As with a data inventory, tools exist to help with this, but a simple spreadsheet can also be used. This spreadsheet should record:

- The source of your data.
- Why it's being collected.
- Each person or team in your organization that touches it.
- Plans for its deletion or retention once the data has served its purpose.[4]

4 **USE CASES**: Use cases put together a hypothetical question the company is trying to solve, which helps define the information needed and the system the data will come from. For example, *How can the organization predict which third-party vendors have the*

"For example, imagine you have to sort through a million files for the word "blue." Even if it only took you one second per file, you'd have to sort for more than 11 days straight without stopping to sleep or eat."

"Once the company conducts the initial data audit, it must decide what data is meaningful for its strategic goals. It is not enough to have lots of data; bad data can cost companies millions of dollars."

same characteristics as those where corruption has been an issue? Machine learning therefore requires strategic planning and invested thought about what the company is trying to solve for in order to guide data points to build an appropriate algorithm.

Before the organization can begin releasing the results for real-time use, the model needs to be tested. Many erroneously believe that the machine learns independently and then returns valuable information. In actuality, the algorithms and use cases are the building blocks of machine learning and, therefore, the very foundation from which the bias springs from. Bias in this phase of developing an algorithm can lead to trouble for every future interaction with the model. It is at this phase that organizations must understand the risks that can be created for a company if one does not fully understand the algorithm and the data used and whether the outcomes can be relied upon. This applies both to the initial process of building algorithms and to the compliance role of evaluating models that others have developed. We cover this further under "Ethical Implications in AI" below.

5 Investment/phasing: Work on solving one or two key risk questions through use cases, then build from there, as the need to build the use cases and to test the data quality and the outcomes for biases takes time.

Once the company conducts the initial data audit, it must decide what data is meaningful for its strategic goals. It is not enough to have lots of data; bad data can cost companies millions of dollars.[5] Even if you have lots of good data, that data may be incomplete (e.g., missing contact numbers or emails). You will also need to scrub or cleanse your data to ensure that duplicates do not exist. As you examine data, remember the company's strategic goals. Think about what data is most meaningful to risk identification or prevention and can help measure the effectiveness of your compliance program.

Connecting the dots across systems to yield desired and reliable insights

AI will interpret and learn from the data you have prepared. Your AI will then progress in flexible, adaptive learning to use data to achieve strategic company goals. Adaptive learning simply means that the algorithms created with the data sets will "learn" and adapt through interaction with

the AI system and the user. "For some companies, they simply would never have the level of rich, insightful data, in the quantities needed to drive quality algorithms. To meet growing demand, insight-as-a-service is emerging as a way to combine structured and unstructured data, external and internal data, into actionable analytics."[6]

After the data audit, the continuing task of the owner, with guidance from the compliance team, will be to understand how data sets and algorithms could lead to biased outputs.[7] Some AI specialists have sought to avoid this trap through randomization of data.[8] It is important that the compliance team understand how programming can unintentionally result in bias.

Practical Applications of AI

The science of big data and the algorithms that can make the data useful is interesting. Let's take a look at the reality of where we are today in utilizing AI to optimize business practices, enhance diagnostic research, predict behavior and all the other amazing applications where we have likely only scratched the surface.

From a compliance officer's view, data protection, preservation and retention are paramount. If one is operating in a regulated environment with data classified as personally identifiable information and its highly regulated sub-categories of payment card information and protected health information, the regulatory compliance team of any organization needs to be viewed as an ally and to be brought into discussions and research on potential uses of AI at the initiation of any project.

Now, expand this data universe globally to internal controls required of personal information processed across borders and governed by the EU's General Data Protection Regulations, the EU Privacy Shield, the Philippines's National Privacy Commission and all the other emerging cross-border transfers of personal data regulations. AI applications and databases must have controls built into their programming on the front end of design to protect and house this data for compliance with all governing regulations.

"From a compliance officer's view, data protection, preservation and retention are paramount."

> "An AI bot can detect voice inflections to discern the quality of the interaction—happy, angry, frustrated—better than most humans can. It can even understand when an interaction has maxed out and needs to be escalated to a human."

Complicating this real-life application of AI is a consumer's "right to be forgotten" under privacy regulations such as the *California Consumer Privacy Act.* How do we teach an AI tool to "tag" this information so the algorithm can search and destroy in all locations when needed?

To date, outside of research applications, AI is predominately applied to streamlining business processes involving interactions between humans. This is typically done with an AI bot or "chatbot" (like a chat robot) that simulates human conversation. Imagine calling your bank or any customer service line and getting three questions into the "conversation" before you realize you are not talking to a human being. The AI bot is responding to you based on millions of past interactions where its database and underlying data analytics infrastructure have analyzed the optimal responses to create the most engaging and pleasing consumer experience. The same applies to such interactions with consumers on platforms such as social media and email. In the compliance space, an AI bot can also be leveraged to help employees search their employer's policy documents and return specific content or links to more information.

An AI bot can detect voice inflections to discern the quality of the interaction—happy, angry, frustrated—better than most humans can. It can even understand when an interaction has maxed out and needs to be escalated to a human. Words can be triggers to human escalation. Imagine a customer service interaction about a consumer product where the caller states "I got sick" or "I had to go to the hospital." The bot recognizes the trigger words "sick" and "hospital" and escalates the call after a quick and appropriate transfer message to keep the caller or social media contact engaged. Such prompts and triggers to get a consumer to the appropriate human professional are imperative in an FDA-governed product world.

Now let's switch to a medical footprint. The old world of medical transcription is being transformed. A physician conducting an exam may no longer have to take the time to type notes into a customer relationship management tool housing a patient's record, and a surgeon may be able to log surgical procedure notes by narrating what he or she is doing during the procedure. Behind the scenes, it may no longer be a human medical transcriptionist translating the voice into text and data and then entering it into a database; rather, it may be an AI bot. The bot utilizes AI to convert voice to text, taking out the human factor entirely, with potentially greater data accuracy and, thus, an optimal patient experience

and medical record to be shared among collaborative caregivers. The AI novice may conclude that bots will replace humans and job availability will decrease. That's not entirely correct. Yes, bots will replace humans in the more basic and common interactions, but they will escalate to a human those longer, more complex interactions, assisting organizations in driving profitability. As we have already learned above, humans must interact and provide the "intelligence" piece during the planning phase, as well as evaluate the results the AI model is yielding, to ensure accuracy and avoid potential bias. With AI as a partner, the speed at which we can process vast amounts of information increases exponentially, and it helps to focus attention in targeted or outlier areas that may help ensure resources are deployed in a truly risk-based manner.

Today, AI is just realizing its potential and return on investment. For the compliance officer, it is essential to be part of the design of any application to provide appropriate guidance on the required internal controls that should be built into the skeletal algorithm to ensure accuracy, protect and preserve data, meet data retention requirements, and avoid bias and uphold your company's standard of ethics. That takes us to the next and final area of this paper: what are the ethical implications of AI that we should consider?

"Yes, bots will replace humans in the more basic and common interactions, but they will escalate to a human those longer, more complex interactions assisting organizations in driving profitability."

Ethical Implications of AI

The table below explains five principles that should be applied when considering the ethical implications of a new AI development. These principles are further explained in the subsequent pages of this section.

Principles	Transparency	Business Strategy	Trust	Privacy and Security	Values and Social Impact
Considerations	Disclosure of how AI will be used and how it works Transparent business practices for fair use (e.g., clear governance, policies on use of surveillance data) Change management aspects of introducing a new AI solution, employee impacts Financial sustainability Forums and oversight for public use impacts (drive discourse on dislocation of markets due to new tech such as autonomous cars)	Aligned to business strategy with senior management buy-in Governance infrastructure including AI ethics committee Brand and reputation considerations Liability and risk mitigation	Make AI trustworthy Recognize and eliminate bias Guard against mistakes Identify inadequate or biased data Design systems, products and services that can be equitably accessed and used by all Corporate inclusion Societal inclusion	Keeping customer information safe Keeping AI safe from adversaries Ensure AI is not used nefariously Resilient and robust Safe for individuals (protection for vulnerable groups such as children, and protection against hate) Safe for society (protection of civil rights and liberties)	Make AI trustworthy Recognize and eliminate bias Guard against mistakes Identify inadequate or biased data Design systems, products and services that can be equitably accessed and used by all Corporate inclusion Societal inclusion
Risks	Unintended consequences commonly associated with AI, including bias and privacy management Industry- or application-specific risks	Recognize probability of unintended consequences Defining intent (individual company strategy vs. strategy for entire sectors and economies) Insufficient change management plan	Lack of diversity in AI workforce Technology development outpacing identification and mitigation of risks Emerging and potentially conflicting standards, regulations and oversight Liability	Bad actors (nation states, dark web, scam artists, political adversaries) Not a matter of if but when (there really is no 100% secure assurance, so planning for adverse outcomes is a must)	Job obsolescence, rapidly evolving labor market

Principles	Transparency	Business Strategy	Trust	Privacy and Security	Values and Social Impact
Impacts	Impacts may be real or perceived, and may be positive or negative Consider employees Consider customers Consider others in industry Consider society	Systems can be fooled Certain groups get all of the benefits	Damage to individuals Risk to organizations Societal impacts Legal issues: laws lag behind technology	Fallout/disaster recovery Legal issues: different regulations in different jurisdictions	Damage to brand, business and ecosystem Risks to marginalized communities
Enablers	Education Communication Locking arms with influencers	Use of AI to identify problems (e.g., gender gaps, discrimination) Proactive public and private regulation	Increase diversity in AI workforce Human and machine learning capabilities Integration of psychology, humanities into plans before building Validation testing Proactive work with regulators and oversight agencies	Proactive work with regulators, law enforcement Deep diving with privacy/security personnel on where we can and cannot take "smart" risk	Define most common values and norms in advance and agree on leveraging those as baseline

"Ensuring that ethical principles are embedded into AI applications and processes at every stage of the lifecycle will enable companies and governments to earn the public's trust in systems."

Ethical Implications Deep Dive

The rapid maturation of AI and its application in widely varying domains bring both tremendous opportunity as well as fear of the unknown. Ensuring that ethical principles are embedded into AI applications and processes at every stage of the lifecycle will enable companies and governments to earn the public's trust in systems. Unfortunately, a number of examples exist where systems have demonstrated biases and discrimination that closely mirror and amplify real-world inequalities and where systems have exhibited serious shortcomings in areas such as data privacy. Potential vulnerabilities include misuse of AI technology for nefarious intent, as well as concerns with transparency and accountability. Therefore, to engender public trust, it is imperative that AI be developed, used and protected in a way that is safe, secure and fair for all. The five principles shared in the table above (Transparency – Business Strategy – Trust – Privacy and Security – Values and Social Impact) provide a model for proper consideration of the ethical implications of AI in its development and application.

1. Transparency

BACKGROUND

While there have been many successful AI implementations that have benefited their organizations, customers, employees and society, there have also been highly publicized stories of AI shortcomings resulting in reputational damage and skepticism about the ability to use this technology fairly, safely and ethically.

CONSIDERATIONS

Here are some important elements to consider:

- **Disclose how AI will be used.** Publish a statement of AI principles and make it available to all stakeholders, internal and external.
- **Disclose how AI works.** Decisions made by AI should be traceable and explainable by humans.
- **Articulate business practices for fair use of AI.** Translate the AI principles into a clear governance structure, including policies for the internal application of AI as well as for the products and services the organization provides. Be specific regarding any unacceptable

applications, such as limitations on the use of AI-generated surveillance data.

- **Ensure robust change management.** This is an essential feature of introducing AI. Plan for and manage the necessary changes throughout the infrastructure and/or product and service structure, considering impacts to all stakeholders and any necessary business process and job re-engineering or obsolescence.
- **Ensure AI implementation is financially sustainable**. Consider what resources will be required annually to maintain and use both the AI system and any of the information it generates. Determine whether the cost-benefit analysis in favor of an initial AI implementation will still hold true through a business downturn.
- **Provide forums and oversight for impact discussions**. For example, there may be a need for public discourse on dislocation of markets due to new technology such as autonomous cars. Or an organization might provide a discussion forum for employees concerned about the use of AI in performance evaluation processes.
- **Provide "mythbusters."** A specific group in the organization should be charged with being the go-to place for questions, concerns and clarifications.

RISKS

- **Application of AI can produce unintended real or perceived consequences.** Biased decisions and compromised privacy have already surfaced as real concerns in consumer and personnel applications. However, some perceived consequences with little basis in reality, such as robots or weapons running amuck, should also be addressed in communications as applications are developed.
- **There may be industry-specific or application-specific considerations**. For example, AI application to personnel systems, weapons systems or commercial product manufacturing may require different transparency practices. This should be addressed as principles, policies and procedures, training, and communications are developed.

IMPACTS

- **Transparency needs differ by stakeholder group and the way AI is being applied.** Implementation of AI in a company's infrastructure impacts people and organizations differently than implementation in a product or for customer service. An organization should ask itself:
  - How will this affect employees? Will jobs be lost? Is retraining an option?

"Biased decisions and compromised privacy have already surfaced as real concerns in consumer and personnel applications."

> "As technology continues to evolve and AI solutions become more prevalent, it's clear there needs to be a solid strategy in place that is fully aligned to the organization's business strategy."

- How will this affect customers? What advantages or opportunities are we creating? Might there be unintended negative perceptions?
- How will this affect others in our industry? Are we fundamentally shifting product or service capability in a way that might drive other innovation or, conversely, that might be seen as anti-competitive?
- How will this affect society? Do we need to explain our testing or the limits we are building into the system?

ENABLERS

Education and communication are key to enabling transparency. This includes identification and delivery of the training needed for all stakeholders—especially those charged with defining and developing AI-enabled systems, but also those who will use the systems. Influencers, both internal and external, can drive the necessary culture shifts required for organizations to successfully embrace the use of AI.

2. Business Strategy

BACKGROUND

As technology continues to evolve and AI solutions become more prevalent, it's clear there needs to be a solid strategy in place that is fully aligned to the organization's business strategy. These solutions must be incorporated into the risk assessment and mitigation processes, and brand and/or reputation impacts must be considered. A governance structure should be codified to ensure the proper stakeholders have a collective understanding of the plan and have strategic input and decision-making throughout all phases of scoping, development, implementation and use.

CONSIDERATIONS

Here are some important elements to consider:

- **Align to and embed within organizational strategy:** AI should be leveraged to advance an organization's strategic agenda or to help fulfill its purpose. In addition, there should be buy-in from the organization's senior management team.
- **Governance infrastructure:** In addition to focusing on data and technology governance, it is advised that a committee (or similar forum) be dedicated to overseeing ethical considerations that may arise throughout the lifecycle or to evaluating escalations (from inside or outside of the organization). Governance bodies should be cross-functional in nature and should drive transparency and organizational accountability.

- **Brand and Reputation:** When considering the direction of an AI solution, identify and, where necessary, plan mitigation strategies for how the solution will be perceived outside of the organization. Will it enhance or hurt the organization's brand or reputation? Consider all potential stakeholders that could impact perception, including investors, NGOs, prospective employees, etc.
- **Risk assessment and mitigation:** Bake AI technology solution plans into risk assessment processes as well as mitigation plans to ensure all aspects of risk have been identified. Use this information to influence decisions and resources and to outline tactics and measurements to ensure "smart" risks are well managed.

RISKS

In addition to risk assessment and risk mitigation, there are risks to the plan and strategy failing. Some examples are:

- **Not recognizing the probability of unintended consequences:** Especially in early stages, AI has a greater chance of false positives as well as incorporating bias without warning. Starting small, conducting extensive pilots and using less risky use cases is a best practice.
- **Not properly defining the intent or setting expectations:** Clearly state whether you are trying to influence an entire sector or just your own organization. The answer will define your risk profile and will influence all aspects of your plan.
- **Insufficient change management plan and follow-through:** Innovation can fail for many reasons, but one of the leading causes is lack of a true change management plan that ensures there is rigor and a process to follow through and drive the change.

IMPACTS

- **Systems can be fooled:** Through machine learning, systems could "learn" wrong or unintended things that ultimately produce outcomes not originally intended. In this scenario, post-deployment information is introduced to the AI system with the intent to "fool" the system to produce an outcome that may have negative impacts or could provide targeted personal benefit. Businesses and industries would need to be mindful of this potential and, at the outset, create a plan to immediately react and reduce risk should the issue occur.
- **Certain groups get all of the benefits:** AI should not be used to benefit only one group at the expense of others or be developed in a way that only powerful segments of society would have access to it or the outcomes. The ultimate business intent should be to design something that has the broadest and most meaningful impact, if the solution is meant to provide benefit outside of the organization.

"Especially in early stages, AI has a greater chance of false positives as well as incorporating bias without warning. Starting small, conducting extensive pilots and using less risky use cases is a best practice."

> "As businesses and governments delve deeper into the development, deployment and use of AI technologies, a critical consideration is how to ensure public trust in these AI systems that will affect virtually every aspect of human life."

ENABLERS

- **AI to identify issues:** AI could be leveraged to help an organization identify gaps or issues of which it was unaware, such as gender gaps or discrimination. The AI solution could help pinpoint negative trends, emerging risks or other helpful data points. AI may then be redirected to help mitigate key issues to resolve those same gaps.
- **Leveraging values-based decision making process or model:** An enabler in the development process and ongoing use of AI would be a comprehensive organizational values-based decision-making model, which the enterprise already leverages and trusts. This will ensure the organization's values are thoughtfully integrated up front, but also when reviewing output, to ensure information is actionable only when it also adheres to the values.

3. Trust

BACKGROUND

As businesses and governments delve deeper into the development, deployment and use of AI technologies, a critical consideration is how to ensure public trust in these AI systems that will affect virtually every aspect of human life. The public must perceive that AI is being developed, used and protected in a way that is safe, secure and fair for all.

CONSIDERATIONS

Many entities are currently reaching the crossroads between experimentation with and full-scale implementation of AI. Responsible adoption of AI must take into account a number of considerations in establishing ethical frameworks for the trustworthy development, deployment and use of AI. To that end, in April 2019 the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) released a set of "Ethics Guidelines for Trustworthy Artificial Intelligence."[10] The guidelines list seven key requirements that AI systems should meet in order to be trustworthy:

1. **Human agency and oversight**: Including fundamental rights, human agency and human oversight;
2. **Technical robustness and safety**: Including resilience to attack and security; fallback plan; and general safety, accuracy, reliability and reproducibility;
3. **Privacy and data governance**: Including respect for privacy, quality and integrity of data, and access to data;
4. **Transparency**: Including traceability, explainability and communication;

5. **Diversity, nondiscrimination and fairness**: Including accessibility and universal design, stakeholder participation, and the avoidance of unfair bias;
6. **Societal and environmental well-being**: Including sustainability and environmental friendliness, social impact, society, and democracy;
7. **Accountability**: Including auditability, minimization and reporting of negative impact, trade-offs and redress.

Our research has found that many organizations, both governmental and private, have developed principles and guidance on these issues. The Defense Innovation Board released principles in October of 2019 that were adopted by the Department of Defense in February 2020.[11] The five principles—Responsible, Equitable, Traceable, Reliable and Governable—apply to both combat and non-combat systems. In addition, the work of the Institute of Electrical and Electronics Engineers (IEEE), a membership association made up of approximately 430,000 engineers, scientists, students, and programmers and developers, is particularly helpful. IEEE has recently published a treatise on this subject, available for free to the public, entitled *Ethically Aligned Design: First Edition* (EAD First Edition) (https://ethicsinaction.ieee.org/).

"Our research has found that many organizations, both governmental and private, have developed principles and guidance on these issues."

EAD First Edition provides a well-thought-out set of eight general principles to consider that can be used as further guidance on how to structure a cross-functional working group. For instance, depending on the industry and regulatory environment in which your organization sits, a master working group could be created with respect to each of the eight principles. Each group could be tasked with identifying the impacts of one of the general principles laid out by EAD and how, or if, those impacts are relevant to the organization.

HUMAN RIGHTS: The Human Rights group could be tasked with ensuring that AI is created and operated to respect, promote and protect internationally recognized human rights.

WELL-BEING: The Well-Being group could consider adopting increased human well-being as a primary success criterion for development.

DATA AGENCY: The Data Agency group could determine how to empower individuals with the ability to access and securely share their data and how to maintain people's capacity to have control over their identity.

> "AI systems have been shown to adopt the implicit biases of human developers. This concern is exacerbated by an extreme lack of diversity in the AI development community."

EFFECTIVENESS: The Effectiveness group could work to provide evidence of the effectiveness and fitness for purpose of AI.

ACCOUNTABILITY: The Accountability group could ensure AI is created and operated to provide an unambiguous rationale for all decisions made by the program.

TRANSPARENCY: The Transparency group could work to ensure that the basis of a particular AI decision is always discoverable.

AWARENESS OF MISUSE: The Awareness of Misuse group would be charged with brainstorming ways to guard against all potential misuse and risks of AI in operation.

COMPETENCE: The Competence group could ensure that those working on the AI projects are specified and adhere to the knowledge and skills required for safe and effective operations of the program.

Because every company's issues will be vastly different, the above examples are intended to be potential frameworks that may be used to begin discussions at your organization.

Risks

There are notable risks to successful implementation of trustworthy AI, including:

- **Implicit bias and lack of diversity in AI development community:** AI systems have been shown to adopt the implicit biases of human developers. This concern is exacerbated by an extreme lack of diversity in the AI development community. According to a recent study, minorities and non-male genders are significantly underrepresented on AI conference agendas, in AI academia and in the AI workforce.[9] It appears that AI systems are replicating patterns of racial and gender inequality in ways that could deepen and justify historical biases and perpetuate existing forms of structural inequality. The commercial deployment of these compromised tools is cause for deep concern.
- **Lack of regulation, legislation, common standards and oversight**: There are currently no established development standards, regulations or oversight agencies and no mandatory validation testing requirements.

Impacts

- **Damage to individuals and marginalized or underrepresented communities.** As described in our bias examples, if test data itself fails to include broader categories (e.g., underrepresentation of certain races or genders in specific roles), basing models and results on such limited data can further exacerbate the situation.
- **Risks to organizations and society (liability, financial/reputational risks, crime and workforce/economic impacts).** Regulatory requirements, complaints, public opinions and societal expectations can potentially drive consequences for companies when limitations of data or bias is a factor in producing false positives or false negatives in your artificial intelligence models (e.g., leveraging these results to guide hiring decisions).

Enablers

Addressing bias and discrimination in AI must be comprehensive and systematic. Simply locating individual biases within a given technical system and attempting to fix them by tweaking algorithms is inadequate and may even be impossible, particularly if the data upon which the technology is built is systemically biased. Not only does the development community need to recognize and guard against bias in its data and technology, it also must strive to achieve broad diversity within the community (gender, sexual orientation, ethnicity and intellectual diversity/diversity of thought) in order to ensure the systems are inclusive and reflective of our communities at large. The field of research on bias and fairness needs to include a wider social and cultural analysis of how AI is used in context.

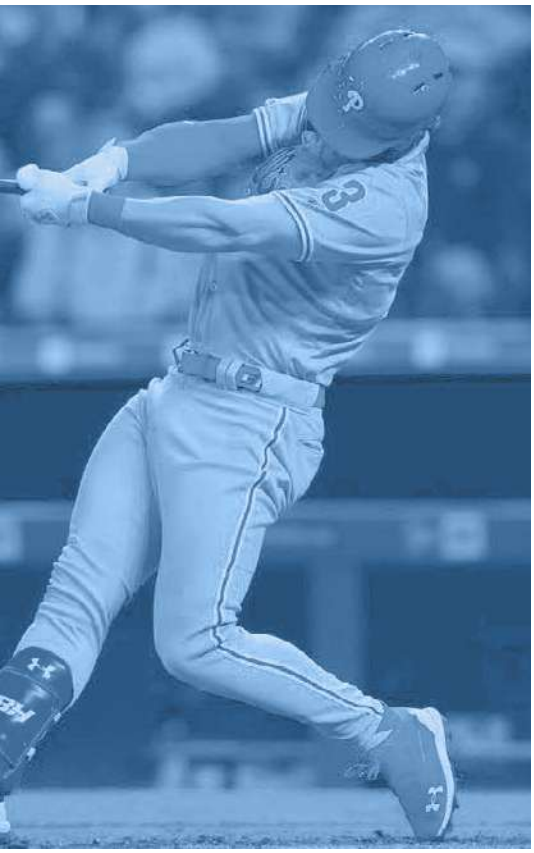
Rigorous development standards and validation testing should be required across the lifecycle of AI systems in sensitive domains. Pre-release trials, independent auditing and ongoing monitoring are necessary to test for bias, discrimination, security and privacy vulnerabilities, and other harms.

To ensure adequate oversight, regulation should be considered for use of AI in sensitive areas such as criminal justice, healthcare and warfare.

Let's explore unintended bias further.

"Simply locating individual biases within a given technical system and attempting to fix them by tweaking algorithms is inadequate and may even be impossible, particularly if the data upon which the technology is built is systemically biased."

> "If the training data input to teach the system what a good hire looks like consists only of resumes of men, the system will spit out resumes of women as not being a good fit for the position."



Case Study: Justice System Algorithms (https://thenextweb.com/artificial-intelligence/2019/02/07/why-the-criminal-justice-system-should-abandon-algorithms/)

Algorithm-based systems can run into various "bias traps." Algorithms are by nature mathematical; they are not based on hunches or experiences. Therefore, social concepts like justice and fairness are not factored into algorithms. This is because justice and fairness are constantly debated and subject to public opinion.

It is in the machine learning phase where bias can appear. The signals, or algorithms, a company creates can have the potential to produce biased outcomes. Erroneous assumptions may be concluded about data sets; therefore, use cases will become an important part of discovering bias at the initial set-up phase. Use cases will test interactions between your information and the system that can help to identify bias before your system is launched.

Consider an algorithm that has been built to determine whether a customer might default on a loan. If the only training data input into the system to teach it what a low-risk loan looks like is data on loans made to white males, the system may unfairly flag loans to people of other races and genders as high risk. Or how about whether a male or a female will be a better hire? If the training data input to teach the system what a good hire looks like consists only of resumes of men, the system will spit out resumes of women as not being a good fit for the position. By now, you are likely able to get a feel for the importance of the training data and how biased, wrong or skewed information can result in unintended consequences.

Privacy and Security

Background

Technology is now enabling us to develop more complex analytical tools that allow for connections and linkages to be made quickly between different data sets while deploying AI models. As organizations make this journey, they need to be fully aware of the privacy and security of the data for customers, employees and others.

Considerations

Here are some important elements to consider:

- **Keeping information safe**: Information held within AI models should be secure and handled in accordance with the organization's cybersecurity policies. Any personally identifiable information from customers, employees or others must be handled in accordance with data privacy and security laws in the jurisdiction where the AI is being deployed.
- **Keeping AI models safe**: What would be the impact if an AI model fell into the wrong hands? AI models should be stored securely in accordance with the organization's cybersecurity policies.
- **Keeping AI safe for individuals**: AI models should be safe for individuals, and consideration should be given to protection for any vulnerable groups (e.g., children) and protection against any forms of discrimination.
- **Keeping AI safe for society**: AI models should be safe for society, and consideration should be given to protect civil rights and liberties.
- **Using AI for the right reasons**: AI should be used for the right reasons and not nefariously.

Risks

As part of the AI model deployment, the data privacy and security risks for the model should be assessed using the organization's risk management approach.

Here are some factors to consider for the risk assessment:

- **Reputational impact**: If the AI model were compromised by a significant bias or at the hands of bad actors, what would be the impact on the organization?

> "Information held within AI models should be secure and handled in accordance with the organization's cybersecurity policies."

- **Data privacy laws**: If the data used in the AI model were compromised, how would this need to be reported and to whom?

Impacts

The reputational and financial impact of a failure in the management of data privacy or cybersecurity of AI could result in a loss of customer base and related revenue or in potential fines from regulators.

Enablers

AI is a new and fast-paced technology practice. As such, it provides compliance professionals the opportunity to help ensure AI is developed with the correct security and privacy requirements for ethical deployment within their organizations. Close engagement with technology departments, data privacy lawyers and cybersecurity teams is essential for this to be effective.

AI development also provides the opportunity to work with regulators to help further define how AI models should and will be deployed, along with the regulation that should support these deployments.

> "It is important to recognize that AI needs to be developed, deployed and used with an ethical purpose. There need to be guiding principles that inform and dictate the process."

Values and Social Impact

Background

It is important to recognize that AI needs to be developed, deployed and used with an ethical purpose. There need to be guiding principles that inform and dictate the process.

One principle to consider is values and societal impact—that is, does the underlying AI align with values (societal, industry, personal, corporate, etc.) and with societal norms, and what is the broader impact on society and the population? *Trustworthy AI should be ethical, ensuring adherence to ethical principles and values throughout the system's entire lifecycle.*[10]

Considerations

The purpose behind AI should be advancement of society; in other words, AI should produce a social benefit. It is important to consider both the long- and short-term implications AI will have socially.

One must recognize the potential negative impacts AI can have on behavior. While AI is transforming efficiency norms, the question to consider is *"Will our increased interactions with AI and machines trigger actions that are not beneficial to our behaviors?"*

Companies must understand and recognize that individuals, employees, consumers and customers may be skeptical or may not trust AI. The purpose of AI should be to foster, produce and create trust among the culture. While there may be certain industries where the number of jobs could be negatively impacted, there will be industries that will experience increases as well as the opportunity for more skilled or complex roles. It will be crucial to realize that there will also be the need for a combination of human and artificial intelligence, and that new roles and responsibilities will need to be identified.

Risks/Impacts

AI has the potential to lead to a rapidly changing or evolving labor market, so companies and individuals will need to be cognizant of and prepare for that. Organizations will need to train and invest in employees to meet the changes.

AI will need to be designed and monitored to ensure that it benefits everyone, not just a defined few, and that individuals are not excluded based on age, social status, gender, etc. AI should not disadvantage people, nor should it produce economic disadvantage to segments of the population.

Negative outcomes or unforeseen circumstances resulting from AI have the potential to damage one's brand and/or reputation, resulting in a loss of employee, consumer or regulator confidence and trust. There is also the potential for a broader behavioral or societal impact with more, long-lasting, irreversible or enduring effects.

Enablers

It will be critical that, at a minimum, organizations identify common values and norms that drive their strategic development and deployment of AI. Companies will need to have some form of ethical framework when it comes to AI.

To the extent possible, ethical behavior requirements will need to be considered, incorporated and built into systems, so ethical training

"Companies must understand and recognize that individuals, employees, consumers and customers may be skeptical or may not trust AI. The purpose of AI should be to foster, produce and create trust among the culture."

for AI developers will need to be considered. There will also need to be a means for providing feedback and monitoring feedback on systems, to recognize the unintended consequences. This will enable developers to make changes or modify the systems where appropriate and, most importantly, to communicate those impacts and steps taken to the broader community.

> "First and foremost, introduce yourself to those doing the actual development work of the algorithms or machine learning products. New relationships that generally have not been nurtured in the corporate environment will be essential."

Conclusion

So what should we do now? How does a person who is becoming acquainted with the topic of AI and its related issues begin to tackle them on a company-wide scale? One step is to enhance your knowledge and skills in order to function effectively as a strategic advisor to your business. Another is to understand where it may be beneficial to incorporate AI into your compliance program strategic priorities.

First and foremost, introduce yourself to those doing the actual development work of the algorithms or machine learning products. New relationships that generally have not been nurtured in the corporate environment will be essential. Product developers, programmers, engineers and data scientists are advised to get to know their compliance, ethics, human resources, IT infrastructure/security, operations, sales, marketing and legal teams—and vice versa. It sounds overwhelming, but the creation of cross-functional teams across the company will be critical to allow for diversity of thought and experiences.

You have now taken your first step in navigating AI to optimize risk management and minimize ethical implications.

FOOTNOTES:

[1] Marr, B. (2018, April 30). 27 Incredible Examples Of AI And Machine Learning In Practice. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice/#4075e79c7502

[2] https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/

[3] Greene, T. (2018, August 2). A beginner's guide to AI: Algorithms. *TNW*. Retrieved from https://thenextweb.com/artificial-intelligence/2018/08/02/a-beginners-guide-to-ai-algorithms/

[4] Making a Personal Data Inventory and Data Map in the GDPR Era. (2019). *Siteimprove*. Retrieved from https://siteimprove.com/en/gdpr/personal-data-inventory-data-maps/

[5] Melendez, C. (2017, December 29). Your company is getting pumped about AI—but is your data ready? *InfoWorld*. Retrieved from https://www.infoworld.com/article/3244474/your-company-is-getting-pumped-about-aibut-is-your-data-ready.html

[6] Ibid

[7] Greene, T. (2019, February). Why the criminal justice system should abandon algorithms. *TNW*. Retrieved from https://thenextweb.com/artificial-intelligence/2019/02/07/why-the-criminal-justice-system-should-abandon-algorithms/

[8] Hester-Reilly, H. (2019, June 4). The Chris Butler Hypothesis: Adversarial Product Management Gets to the Core of What Really Matters Using Contrarian Thinking. *H2R Product Science*. Retrieved from https://h2rproductscience.com/ep018-chris-butler/

[9] West, S.M., Whittaker, M., and Crawford, K. (2019). *DISCRIMINATING SYSTEMS Gender, Race, and Power in AI* [PDF file]. AI Now Institute. Retrieved from https://ainowinstitute.org/discriminatingsystems.pdf

[10] European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[11] https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Ethics & Compliance Initiative
8409 Lee Highway, #2175
Merrifield, VA 22116-8239

ETHICS.ORG

ISBN: 979-8-3507-3376-1

ECI