TOPIC: RESPONSIBLE AI

Contributors:

Andre Schwan – Gijima
Harish Mekerira - Amazon
Mark Williams – TachTech
Matt Estes - Disney
Ryan Skipp—T-Systems
Shamir Charania – JOT Digital
Tom Scott—ADP

CONTENTS

TABLE OF CONTENTS

LEGAL NOTICE

LEGAL DISCLAIMER:

THIS DOCUMENT AND THE INFORMATION CONTAINED HEREIN IS PROVIDED ON AN "**AS IS**" BASIS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, THE ALLIANCE (ALONG WITH THE CONTRIBUTORS TO THIS DOCUMENT) HEREBY DISCLAIM ALL REPRESENTATIONS, WARRANTIES AND/OR COVENANTS, EITHER EXPRESS OR IMPLIED, STATUTORY OR AT COMMON LAW, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, VALIDITY, AND/OR NONINFRINGEMENT. THE INFORMATION CONTAINED IN THIS DOCUMENT IS FOR INFORMATIONAL PURPOSES ONLY AND THE ALLIANCE MAKES NO REPRESENTATIONS, WARRANTIES AND/OR COVENANTS AS TO THE RESULTS THAT MAY BE OBTAINED FROM THE USE OF, OR RELIANCE ON, ANY INFORMATION SET FORTH IN THIS DOCUMENT, OR AS TO THE ACCURACY OR RELIABILITY OF SUCH INFORMATION. EXCEPT AS OTHERWISE EXPRESSLY SET FORTH HEREIN, NOTHING CONTAINED IN THIS DOCUMENT SHALL BE DEEMED AS GRANTING YOU ANY KIND OF LICENSE IN THE DOCUMENT, OR ANY OF ITS CONTENTS, EITHER EXPRESSLY OR IMPLIEDLY, OR TO ANY INTELLECTUAL PROPERTY OWNED OR CONTROLLED BY THE ALLIANCE, INCLUDING, WITHOUT LIMITATION, ANY TRADEMARKS OF THE ALLIANCE.

TRADEMARKS: OPEN ALLIANCE FOR CLOUD ADOPTION^{SM}, OACA^{SM}, and the OPEN ALLIANCE FOR CLOUD ADOPTION logo$^{®}$ are trade names, trademarks, and/or service marks (collectively "**Marks**") owned by Open Alliance for Cloud Adoption - A Linux Foundation Project, Inc. and all rights are reserved therein. Unauthorized use is strictly prohibited. This document does not grant any user of this document any rights to use any of the OACA's Marks. All other service marks, trademarks and trade names reference herein are those of their respective owners.

If any derivatives of any Open Alliance for Cloud Adoption - A Linux Foundation Project, Inc.'s documents are published, the following statement must be identified: These documents are based on original documents created by the Open Alliance for Cloud Adoption - A Linux Foundation Project, Inc. (OACA), but may contain changes to the original OACA document which have not been reviewed or approved by the OACA.

EXECUTIVE SUMMARY

Artificial Intelligence (AI) has become a prominent and transformative technology in many domains, revolutionizing the way we live and work. From a philosophical perspective, AI raises intriguing questions about the nature of intelligence, consciousness, and the relationship between humans and machines. It intersects with fields such as philosophy of mind, cognitive science, and psychology, prompting us to explore the boundaries of human cognition and the potential of machines to emulate human intelligence. Our use of AI does not only have positive effects, however. Attributes of AI systems, such as speed-to-answer and the convenience-of-use, can confuse humans into trusting outputs that are not always correct, or not valid in a given context. Bad actors may also arise, infiltrating or using AI to influence or undermine legitimate objectives. A lack of transparency and broad understanding of AI systems has led to an assumption that AI systems work well in all settings. It is this sometimes-misplaced trust coupled with the inability to determine the validity of the AI's outcomes that introduces risk which organizations need to guard against.

At the OACA, we believe that responsible AI use starts with an understanding that AI technologies are good for human augmentation and not so good for human replacement. The reality is that much of an organization's intellectual property resides within its people, and AI can assist organizations in enabling people to add value further up the value chain. This paper discusses the OACA view that a responsible AI Framework consists of stated ethics, processes, and governance and must be defined for each organization considering various stakeholders' rules, guidelines, and values. The OACA further identifies that this framework must include ethical considerations at the societal level, factoring in relevant legislation and guidelines from applicable governments, vendors of AI technology, and applicable advisory entities. When developing this framework, organizations must consider AI system attributes such as safety, diversity, fairness and equity, transparency, human oversight, validity and robustness, and accountability as discussed in this paper.

Given the dynamic and evolving nature of AI systems, the framework must specifically address continuous learning and improvement processes. These processes typically involve updating models with new data, measuring the distribution of outcomes, and enhancing (or terminating) their performance over time. As AI capabilities continue to advance, it is crucial to continually examine and weigh their implications and impact on society and business against the opportunities it may bring. This includes mapping the potential ethical, social, and legal ramifications of AI use, as well as the responsibilities that come with its deployment. Even with upcoming legislation around AI use, organizations play a significant role in shaping the development and use of AI, and it is essential for them to adopt responsible practices to ensure the ethical and beneficial application of AI technologies.

This responsible AI framework forms the basis for any organization when determining appropriate use cases for the use of AI and the guardrails that must be applied when deploying, integrating, and incorporating AI systems at people, process, and technology levels. By considering these elements and preparing in advance, organizations can effectively and safely leverage AI technologies and processes to drive innovation and achieve their goals.

INTRODUCTION

In the realm of philosophy, AI is often defined as the ability of machines to perform tasks that would require human intelligence. This definition includes capabilities such as understanding natural language, learning from data, option selection, problem-solving, and even exhibiting creative deductions. AI systems aim to replicate or simulate human cognitive processes, enabling machines to perceive, comprehend, and interact with the world in ways that were once exclusive to humans. These elements include AI algorithms and models (ranging from Language Models through Neural Networks, Discriminant Analysis, and Learning vectors, amongst many others), data collection and preprocessing, machine learning techniques, training and evaluation, natural language processing (NLP), computer vision, and deep learning. Importantly, AI systems can generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. (Reference NIST https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf).

In this paper, we focus on AI responsibility within organizations. We discuss the idealized target future state and the recommended actions that organizations should take to bridge the gap. Our analysis will revolve around three main areas: people, process, and technology.

The first area, people, explores the literacy, skills, and talent required within organizations to effectively and responsibly use AI. We delve into the importance of fostering a diverse and inclusive workforce, promoting AI literacy, and ensuring ethical decision-making in AI-related tasks.

The second area, process, examines the activities and practices that organizations should adopt to ensure responsible AI use. These include establishing clear governance frameworks, implementing robust data privacy and security measures, and incorporating ethical considerations throughout the AI lifecycle.

The third area, technology, focuses on the tools and technologies organizations are utilizing to perform AI activities. We explore the advancements in AI algorithms, the importance of explainability and interpretability in AI systems, and the need for ongoing research and development to address emerging challenges.

By addressing these three areas comprehensively, organizations can navigate the complex landscape of AI responsibility and contribute to the development of a responsible and beneficial AI ecosystem.

PEOPLE

In the realm of AI responsibility within organizations, the role of people is paramount. The individuals within an organization need to possess the skills and talent necessary to use AI responsibly. It is essential to acknowledge that even the most advanced AI models, such as those narrowly trained within a specific domain of knowledge, have limitations that only humans can bridge.

One of the key challenges organizations face is ensuring that their people have psychological safety and the necessary skills to navigate the ethical considerations surrounding AI. Training and upskilling programs are crucial to equip employees with the knowledge and understanding of responsible AI use. Organizations can foster a culture of responsible AI adoption by providing their people with the knowledge, tools, and resources to make informed decisions and evaluate/simulate the results of those decisions.

Bias is one such challenge, which can be introduced through the design and implementation of AI algorithms and the selection of data on which they are trained. If the algorithms are not carefully designed and tested for fairness, they can inadvertently discriminate against certain groups or individuals. Famously, facial recognition systems have been found to have higher error rates for people with darker skin tones, leading to biased outcomes and potential harm. On the surface, it may seem that bias is the only reason for these types of outcomes, but there may also be technical implementation reasons. Organizations must have skillsets ready to fully understand the causes of AI outcomes/decisions. Their staff must be empowered to be curious about AI outcomes and constantly monitor outputs against known and emergent forms of bias.

Bias in AI systems is generally considered undesirable as it can perpetuate discrimination and unfair outcomes. However, controlled bias may be necessary in certain situations to address under-representation, ensure fairness, or meet specific domain requirements. The key is to differentiate between harmful, intentional, and unintentional bias, with the ultimate goal being to eliminate harmful biases while responsibly managing beneficial biases. Different applications of AI necessitate different "mixes" of bias to be applied, whether it be data selection or model selection. Diverse teams are crucial for developing responsible and fair AI systems that avoid perpetuating inequality or discrimination. Ultimately, AI systems should be designed with transparency in mind, as transparency can lead to better trust in the AI system outcomes.

People play a vital role in ensuring the ethical use of AI within organizations. At the start of the AI roll-out process, teams are responsible for identifying valid use cases for AI within the organization. Their responsibility includes identifying and addressing potential biases, ensuring transparency in AI processes, and upholding accountability as the AI systems are developed. By actively participating in the development and deployment of AI systems, teams can contribute and learn about the responsible use of AI.

Lastly, other areas of the organization, such as legal, technical, business process, etc., need to gain additional AI responsibility. Everyone will eventually use the AI, and therefore, responsible use is everyone's responsibility! It is crucial for organizations to realize this as they identify and validate AI usecases, customize solutions, mitigate risks, foster innovation, and future-proof the organization. Their knowledge helps optimize internal processes, identify market opportunities, and build trust with stakeholders. By creatively applying AI technologies, they contribute to competitive advantages and ensure the organization stays relevant in the rapidly evolving AI landscape.

PROCESS

Strong processes are vital to an organization's ability to responsibly use AI by helping to navigate the challenges and complexities associated with all aspects of AI use. At the outset, organizations must ensure that AI exploration and roll-out align with their strategic business goals and drivers.

Exploration involves including key stakeholders in cross-functional collaboration to gather diverse perspectives and ideas. By prioritizing opportunities that align with their strategic goals, organizations can focus on high-impact areas where AI can provide significant value.

A thorough data audit and readiness assessment are essential to ensure that the organization's data assets and infrastructure are of sufficient quality, availability, and accessibility to support AI initiatives effectively. Data quality needs to be carefully assessed. One can't mindlessly train an LLM on all enterprise data. It will contain out-of-date, incorrect, or conflicting data. Data leakage is also a concern. Sensitive or proprietary data must be protected. Sometimes, this data is stored in a "convenient" repository rather than one appropriate to the nature and sensitivity of the data. Models trained on enterprise data need to respect authorization controls. For example, one might train a model on HR data, but only people with the same credentials as in the HRIS should be able to query and access the data through an AI-enabled employee chatbot.

If necessary, implementing data governance and data management practices can enhance data readiness for AI use. For example, a data audit may identify critical information that should not be shared or used as an AI model. Strong organizational processes can ensure that required activities are done at appropriate times within the AI system's development lifecycle, which can allow for frictionless exploration while building the foundation for successful adoption.

At a high level, organizations need to have policies and processes in place to govern the transparency, accountability, and fairness of the AI systems in use. Organizations should strive to be transparent about their AI systems, ensuring stakeholders clearly understand how AI is being used and the potential implications. This transparency includes providing explanations of AI-generated outputs, making efforts to demystify the decision-making processes of the underlying AI models, and citing the source data. From an accountability perspective, organizations should establish mechanisms to hold themselves accountable for the outcomes of AI systems. Publicly demonstrating accountability is one way that companies can achieve and maintain credibility in their AI processes. These mechanisms include implementing robust governance frameworks, conducting regular audits, ensuring clear lines of responsibility and oversight for AI-related decisions, and providing disclosure/transparency for AI-influenced outcomes. Tracking of "context drift" over time on original processing must form a part of the governance review process. Fairness is a fundamental principle that organizations should prioritize in their AI processes. It is essential to address harmful biases that may be present in AI models and data, as these biases can lead to discriminatory outcomes. Organizations should invest in bias detection and mitigation techniques, as well as regularly evaluate and monitor the fairness of their AI systems.

Risk management is one such way to ensure transparency, accountability, and fairness in AI use within organizations. Organizations should proactively identify and assess potential risks associated with AI deployment, including privacy concerns, security threats, and the potential for misinformation or disinformation. There are various properties of AI systems that necessitate different risk management approaches than traditional software or systems engineering. For example, the privacy risk of AI search systems could be increased due to the ability of the system to link concepts together in unique and unpredictable ways. Another example is the difficulty in determining how to validate the correctness of AI systems. Traditional software testing strategies may not be adequate given the scale of AI systems (from a decision point perspective) or given the inability to predict/detect side effects of AI systems beyond statistical measures. Risk Management needs to be especially vigilant for bad or false outputs and AI-created outputs as assumptions of truth (hallucinations), coupled with the burden of proof.

Risk management processes for AI systems need to be flexible by design to take into account the constantly and rapidly evolving AI landscape. AI systems' outcomes can be context- and interaction-dependent and require tighter feedback loops between the measurement and governance. Technical and legal organizations need to come together to review overall risk and AI outcomes against stated goals and organizational risk tolerances.

Organizations need to be vigilant to protect the intermediate and final outputs of their AI system and actively monitor and prevent misuse.

AI risk management frameworks have the opportunity to manage not only risk but also positive outcomes of the AI models. By implementing robust risk management practices, specifically for AI-driven processes and outcomes, organizations can mitigate these risks and ensure the responsible use of AI.

Many organizations are turning to third-party solutions to meet their AI needs. Similar to processes required when investigating cloud usage, organizations need to understand the shared responsibility model of these third-party solutions, from the training data, the model, and the original algorithm or model source to the final outcome. Legal teams play a crucial role, as they are responsible for executing contract and indemnification processes and establishing rules and guidelines for using specific AI technology:

- The organization's data capabilities are responsible for evaluating data-related aspects concerning data used for training and production work, such as privacy, security, and integration capabilities.
- The privacy capabilities focus on assessing privacy policies and data handling practices.
- The IT/security capabilities evaluate the security measures implemented by the vendor.
- The procurement capability manages the procurement process and negotiates contracts to ensure end-to-end transparency and responsibility.
- Stakeholders and business units provide input based on their requirements.

By involving these capabilities, organizations can ensure that the tools/software align with legal, data, privacy, and security requirements, mitigating risks and facilitating successful integration into the AI ecosystem.

TECHNOLOGY

In the realm of AI responsibility within organizations, the technology domain plays a crucial role in enabling responsible AI use. It encompasses the tools, technologies, and platforms organizations utilize to perform AI activities while ensuring ethical considerations are met.

Technology to support AI necessarily builds upon technology capabilities already within the organization. For example, authorized use of AI systems requires that all access to the AI system be authenticated, which can be satisfied by existing enterprise capabilities around identity and access management. Another example centers around extending the use of existing input/output sanitization capabilities to protect the inputs and outputs of AI systems. There are other attributes of AI systems, however, that necessitate the introduction of new technology or technology paradigms.

Explainable AI refers to the ability of an AI system to provide understandable explanations and justifications for its decisions and actions, and it is one such area where technology can enhance the current processes. Model architecture can be designed to include explainability features, such as attention mechanisms or interpretable weights, to make the decision-making processes more transparent. Specific auditing and monitoring tools can be incorporated into AI systems to audit and capture information about the inputs, intermediate steps, and outputs of AI systems. These audits can then be used within organizations to explore for biases, errors, or inconsistencies and enable continuous monitoring of the AI system's performance explainability. Lastly, AI systems can be built with visualizations that aim to better communicate the inner workings of the AI system, which can lead to a better understanding of how the AI system arrives at its conclusions, identifies patterns, and produces ostensible insight.

Another key consideration in the technology domain is the potential for AI models to generate misinformation and inaccurate or even harmful outputs. The phenomenon of inaccurate outputs generated by text-generating large language models has been widely documented. Errors in general are a large concern, particularly with AI systems that are designed to run autonomously. Guardrails can be an effective way of pursuing error-free execution of AI systems. For example, guardrails can be placed at the input/output stages of AI systems to determine/govern the validity of responses. Guardrails can also be used to limit the scope/use of AI systems to only intended/approved use cases, limiting chances of misuse. In addition, continuous monitoring and human-controlled and validated learning loops can enhance the accuracy of AI systems.

As with all applications, AI systems need extensive security controls to ensure appropriate usage. Not only can AI models be legitimately used for nefarious purposes, but they can also be misused to gain access to intellectual property used to train and build the model or be subjected to model/data poisoning to produce inaccurate decisions. Organizations should be adapting and extending traditional security capabilities (such as access controls, encryption, firewalls, supply chain inspection, etc.) to their AI systems. To address the growing concerns around the cybersecurity of AI systems, numerous security frameworks are currently being developed. These frameworks aim to establish guidelines and best practices for organizations to safeguard their AI systems against potential threats and vulnerabilities. By implementing these frameworks, organizations can bolster their defenses and mitigate risks associated with the deployment and operation of AI technologies, ultimately fostering a more secure and trustworthy AI ecosystem.

Limiting inputs to AI systems is crucial to safeguard against potential vulnerabilities and threats. By carefully controlling and filtering the data inputs, organizations can reduce the risk of malicious actors exploiting the system through manipulated or compromised data. Data profiling and quality technology can play a vital role in this space by analyzing and assessing the data's quality, integrity, and reliability before being fed into the AI system. These technologies can identify and flag suspicious or anomalous data patterns, ensuring that only clean and trustworthy data is used for training and inference. By limiting access to the AI system and implementing stringent data validation techniques, organizations can enhance the system's robustness, accuracy, and overall security posture. This approach not only helps protect against potential attacks but also ensures the AI system produces reliable and dependable outcomes, instilling confidence in its users and stakeholders.

With sustainable AI, organizations must account for the potential benefits and the associated costs of running AI systems. These systems can be resource-intensive, requiring substantial computational power and storage capabilities. Organizations should consider using technology to help measure the impact of AI systems throughout the AI lifecycle. Several techniques can be considered to reduce associated costs, such as using cloud computing and federated learning techniques that allow AI models to be trained collaboratively within industry verticals.

On the security front, organizations need to consider economic denial of service/sustainability (EDoS) attacks, where malicious actors intentionally exploit vulnerabilities in AI systems to disrupt their functionality or oversubscribe their resources, leading to significant financial losses. Addressing this threat requires organizations to incorporate EDoS attacks into their threat models and develop appropriate mitigation strategies. For example, a hacker who has breached an enterprise's firewall might previously have exfiltrated data for offline analysis. With AI systems, such a hacker could learn a lot about an organization by interrogating an AI-based chatbot that has been trained on proprietary enterprise data. That would make their job of finding valuable data much easier.

Addressing bias in AI systems has become a critical area of focus to promote fairness and inclusivity. To mitigate bias, innovative solutions are emerging to tackle this challenge. One such innovation is the development of bias detection and mitigation tools. These tools leverage advanced algorithms and machine learning techniques to identify and quantify biases within AI models and datasets. Additionally, researchers are exploring techniques like adversarial learning, which introduces counterfactual examples to train AI systems to be more resilient against biased outcomes. Another approach involves augmenting training data with diverse and representative samples to reduce bias. These technology innovations are crucial steps towards creating more equitable and unbiased AI systems that can be trusted by individuals and organizations alike.

Emerging technologies, like confidential computing, hold great promise in making AI systems safer. Confidential computing protects sensitive data and algorithms by safeguarding them even from the infrastructure on which they run. By leveraging hardware-based security mechanisms, such as secure enclaves, confidential computing ensures that data remains encrypted and isolated, even from the cloud service providers themselves. This technology can significantly enhance the security of AI systems, as it mitigates the risks of unauthorized access, tampering, or data breaches. By utilizing confidential computing, organizations can develop and train AI models without exposing proprietary or sensitive information, reducing the potential for intellectual property theft or privacy breaches. This promotes trust and encourages collaboration and data sharing among stakeholders, leading to more robust and safer AI systems overall.

When considering AI system implementations, organizations should carefully evaluate the option of using private deployments or those that leverage their existing partner ecosystem and facilitate the application of appropriate security controls. Private implementations provide customers with the advantage of maintaining full control over their AI infrastructure and data. By leveraging their existing partner ecosystem, teams can tap into the expertise and support of trusted vendors who understand their unique business needs. It is crucial to prioritize security controls in all private and public implementations.

CONCLUSION

In conclusion, the rapid evolution and increasing adoption of AI technologies necessitate the urgent need for responsible AI practices. As AI becomes more pervasive in our society, it is crucial that organizations prioritize ethical considerations, data privacy, and the overall societal impact of AI systems. Achieving responsible AI requires a multi-faceted approach that encompasses people, process, and technology actions.

On the people front, organizations must foster a culture of ethics and accountability, ensuring that individuals involved in AI development and deployment know the potential risks and biases associated with AI systems. This involves providing proper training and education on responsible AI practices and encouraging diversity and inclusivity within AI teams.

Regarding processes, organizations should establish robust governance frameworks and guidelines that address ethical concerns, data privacy, and transparency. Implementing rigorous data management practices, conducting regular audits, and incorporating bias detection and mitigation mechanisms are essential steps towards responsible AI.

In terms of technology, organizations should evaluate emerging technologies such as confidential computing, federated learning, and explainable AI to enhance the safety, privacy, and interpretability of AI systems. Embracing sustainable AI technologies and minimizing the environmental impact of AI operations are also critical aspects of responsible AI.

Achieving responsible AI cannot be accomplished by organizations alone. Given AI's complex and evolving nature, collaboration among stakeholders is imperative. Governments, industry leaders, researchers, and civil society organizations must work hand-in-hand to establish regulatory frameworks, standards, and best practices that promote responsible AI. This collaboration should address AI's ethical, legal, and social implications while fostering innovation and economic growth.

In this rapidly changing AI landscape, responsible AI practices are not a one-time implementation but an ongoing journey. Organizations must stay abreast of the latest developments in AI technologies, continuously reassess their AI systems, and adapt their practices accordingly. By embracing responsible AI and fostering collaboration among stakeholders, organizations can collectively harness the power of AI while ensuring its benefits are realized ethically, responsibly, and sustainably.

Appendix (Current – Target – Action)

Bridging the gap between the current and the target state of responsible AI requires organizations to address the three key domains: people, process, and technology. While all domains are important, starting the journey at the people domain can lay the foundation for creating responsible AI systems.

In the people domain, organizations should prioritize building a culture of ethics, accountability, and diversity. This involves training and educating employees on responsible AI practices, and ensuring they understand the potential risks and biases associated with AI systems. By fostering an environment that values diverse perspectives, organizations can mitigate the risk of biased algorithms and promote fairness in AI decision-making.

Simultaneously, organizations must address the process domain by establishing robust governance frameworks and guidelines for responsible AI. This means defining clear policies, procedures, and accountability mechanisms to ensure ethical considerations are integrated into every stage of AI development and deployment. Regular audits and assessments can help identify and rectify potential ethical issues, while mechanisms for bias detection and mitigation should be implemented to ensure fairness and inclusivity.

In the technology domain, organizations should consider leveraging emerging solutions for developing responsible AI systems. This includes utilizing explainable AI techniques, where AI models provide transparent insights into their decision-making process. Organizations should also prioritize privacy-preserving technologies, such as differential privacy and secure multi-party computation, to protect sensitive data and ensure user privacy.

While starting the journey at the people domain is crucial, organizations must also address the process and technology domains in parallel. Establishing responsible AI practices requires an iterative approach, where continuous evaluation, improvement, and adaptation of processes and technologies are necessary to align with evolving ethical standards and societal needs. This iterative approach is also crucial in helping organizations to create and refine guidelines, policies, reference architectures and frameworks to guide the business on responsible AI adoption and use.

Ultimately, bridging the gap between the current and the target state of responsible AI requires a holistic approach encompassesing the people, process, and technology domains. By prioritizing the people domain, organizations can create an ethical foundation, ensuring responsible AI practices are ingrained in their culture. This sets the stage for developing robust processes and leveraging appropriate technologies that align with ethical considerations, leading to the creation of responsible AI systems that benefit both organizations and society as a whole.

PEOPLE

	Current	**Target**	**Action**
PEOPLE	Significantly hyped expectations and concerns exist around AI generally. Investment is flowing towards AI, often based on “hope” or intangible expectations. AI has climbed up the priority list rapidly as businesses are afraid that they may be missing opportunities that their competitors are taking advantage of, with perceived industry leaders seemingly releasing wonderful new functions based on AI and receiving complimentary news coverage for it, with “free brand visibility increases.”	A clear role exists for people in the organization to be in control of AI services and solutions, and who carry the institutional knowledge of the organization also supported by well-defined AI roles such as Chief of AI, AI Legal representative, etc. It is recognized that one is not replaced by AI, but by a human who uses AI.	Evolve a breadth and depth of AI knowledge within the organization to enable prudent and effective integration into the overarching organizational strategy: Define new roles & responsibilities in conjunction with preparation to deal with appropriate innovation and business opportunities. Cross-negotiate within the organization regarding potential people changes and displacement impact (and resulting bias that may negatively influence AI implementations) - be sure to move the relevant people into AI oversight roles so as to minimize risk of loss of institutional knowledge and leverage them to guide the AI evolution.
	Some roles are already being retired / replaced, (possibly pre-emptively) based on AI’s potential in some businesses.		

	Current	Target	Action
PEOPLE	Skill exists at technical levels, and expectation exists at business level, but governance and process are still very “lite.”	Trained developers, business, and legal people are in place so as to ensure that AI implementations and/or use are understood, sustainable, and defendable. (Ignorance is no excuse.)	Train and inform users and consumers about the correct and appropriate use of AI in product development (which means developing organizational skills (awareness & experience) for those impacted/benefitted).
	There is some distrust of available data and how the algorithms use it (or which it is based on) to influence models and outcomes.	Defined “liability” considerations exist – e.g., relating to missed data & facts that skew outputs now and in the future. E.g., a Developer / writer may carry liability and need to be included in the contractual framework of the service. (Know the source that generates the outcome).	
	Accidental / non-deliberate bias exists; some is beneficial, some is not beneficial, but it is generally not a well understood dimension, together with its risk implications.		
	Businesses are not yet skilled enough to identify non-sensical outcomes – training is needed to help them identify anomalies in a structured manner.	Defined outcomes and associated “Ethics” are published and understood widely in the organization for teams to consider when participating in an AI-related project.	
	People see opportunities but don’t have the supporting expertise in AI in place to help exploit them quickly. People are using AI to help sift through masses of overwhelming data, but not exploiting the full business potential that it brings (due to narrow objectives for analysis or not knowing how to use AI to validate their intuitions and support innovation) People in many industry sectors are not aware of appropriate use cases where AI can assist them in exploiting opportunities and increasing business value.	The organization has developed a deliberate AI integration strategy including tailored awareness training for stakeholders and users.	

PROCESS

	Current	Target	Action
PROCESS	Tangible and intangible high-value business opportunities exist, but are generally only partially exploited or understood, often limited to the innovation functions in organizations.	A framework for managing AI in Innovations is defined, considers risks, for example, describing what data can be used, what algorithms can be used, what keys and access can be used, expected and unexpected outcomes, and guidelines for incorporating ethics. This framework should also address appropriate assurance / insurance coverage.	Perform sandbox testing of use cases (in a safe environment) – down-select relevant use cases, Validate applicability and business potential. Then run fast proof-of-concept projects.
	Some organizations are blocking AI due to a lack of trust in the reliability of outcomes, misunderstood potential value to the company, and perceived risk of reputational damage if something goes wrong.	A framework of transparency exists on how, where, and when AI is used, and how assertations should arise from selected training and data.	Select and communicate which agreed algorithms & testing methods may/should be used in projects and how they should be evolved. Define how "Transparency" will be provided (scrutability / inscrutability), especially where this is part of value chains. Define parameters for achieving input and output data quality – relevance & validation criteria.
	Legislation and ethics frameworks have not yet crystalized sufficiently to provide useful guardrails, nor are there strong deterrents to "bad AI behavior" yet.	A well-defined adoption framework exists with controls and guidelines / limits as to how far the responsibility of AI may reach.	Establish a technology and/or process framework with defined ethics and policies for reviewing necessary compliance to the procuring organization's AI standards.
	Businesses are buying AI capability from technology partners, then evaluating in arrears to try and identify / capture opportunities they may otherwise miss and risk losing to the opposition.	Identified major use cases are defined for the business – (e.g., Use AI vs. develop AI-based solutions), with a supporting framework for selecting the appropriate AI and models to use in order to support/enable the use cases. For example, in a health-related organization, consider selecting certain trusted technologies and standards, training data, and trusted or certified algorithms.	Operationalize end-to-end business processes to support the inclusion of AI tools (decision-making process, legal, Engineering, Security, business operations & logic)

	Current	**Target**	**Action**
PROCESS	Legal, Data, and Privacy roles are not engaged in AI adoption or integration in a coordinated framework, in general. This disconnect means that common terminology and meaning are not yet aligned.	A framework is in place to allocate and support responsibility and governance relating to AI use, and if it does not exist within the organization, then it is contracted from outside the organization – e.g., legal or security experts from third parties.	Define governance and controls for the implementation of AI: Establish AI-related governance to create frameworks for enabling and supporting desirable business evolution through using AI. This governance includes setting ethical guardrails (both mandatory & optional constraints), defining liability, establishing security, compliance & contract frameworks, defining desired outcomes measurement against defined baselines, and establishing quality controls against training data, algorithms, and outcomes. Perform impact assessments of proposed AI-driven services and products. These assessments must also consider the factor of multiple linked AI systems with intersecting risks and impact on their explainability.
	Sourcing, control, and protection of training data and algorithms is “lite” - sources are not well understood or contractually secured in most cases by the consuming organizations or their clients.	Certification bodies review and audit AI products and algorithms that are released for public consumption. (e.g., developers of algorithms must adhere to appropriate codes of conduct in their development and supply the appropriate statements with their work.)	Define Internally-approved algorithms & datasets for training until certification bodies exist for industry sectors (framework of approval including validity period of the output, lifecycle, input-output measurement, prompt use policy, source of algorithm, reversibility potential of outcome to source (loss of internal IP), risk implications, bias with negative or inappropriate impact).

	Current	Target	Action
PROCESS	AI is rapidly changing the business dynamic for competitiveness. Governance functions such as security, legal, and compliance that might be impacted by AI are not yet well understood.	The AI methodology applied in product development, coupled with transparency, accentuates the explainability of an output with method, bill of materials, logic, and data source identification.)	The business evolution must be managed to move from a traditional development lifecycle, where product evolution tends to be pulled along, to AI identifying and driving opportunities and accelerating “from behind.” Examine the company’s ability to respond to suggested evolution strategies since typical (reactive) cost, time, and quality factors are replaced with different factors that AI introduces such as fit, power, and market positioning (time = leader/fast follower/follower). NOTE: Build “reversibility” into the process to remove AI from the system or service later if inappropriate. Tag all AI-generated content.
	Consulting organizations are pushing enterprises to adopt AI quickly, and hyper-scalers are competing to deliver and fulfill early business cases.	The organization can comprehensively evaluate solutions for use, both deeply and broadly, to evaluate the best choice for competitive outcomes.	Understand what AI vendors are supplying precisely. The existing software bill-of-materials needs to be augmented to track how the original training was performed and how the implementation evolved from the original source to business use. This understanding must also extend to the licensing and liability chain accompanying it.
	There is often substantial inefficiency within organizations not using AI, resulting in increased costs and missed opportunities.	Opportunities exist for teams to leverage AI to identify opportunities for improvement in their business and operational environments.	Organizations should include AI in their short and long-term budgeting to ensure solutions are properly resourced and align with the overarching business strategy/objectives.
	There is often a lack of sufficient permission provisioning by data owners or sources.	All data has the required permissions from the data owner before being ingested or used in an AI solution.	Create metadata for all collected data defining source, permissions for use, and data owners.

TECHNOLOGY

	Current	Target	Action
TECHNOLOGY	Practical, achievable, and mature use cases appear to be taking deeper hold in a few general areas. The associated technology, tools, and models have developed substantial early maturity and capability. For example: a. Enterprise IT is leveraging AI for internal operational purposes, such as processing large data volumes like SIEM, b. Developers are using machine programming tools to automate coding tasks, improving productivity, c. Call Centers matching potential products and clients d. Identification of marketing opportunities (pursuing "least business change" opportunities) e. Media development and production	Measures are established and monitored, and AI outputs are aligned to expectations based on defined training and certified approved (unbiased) input data.	Select an AI Hosting or Management Solution that establishes a registry or library of models, learning methods, and approved data sets for training, with the capability to track key transactions or use cases. (e.g., SageMaker) Establish controls/mechanisms to ensure that AI algorithms, data, and outcomes are protected against misuse (internally or externally).
	Very little model/algorithm protection is in place, and sources are not well-known or understood.	Sufficient technical guardrails coupled with security, version release management, and monitoring controls are in place to ensure protection throughout the development and operational phases. These are based on security frameworks and industry best practices.	Select, adapt, and deploy a security framework for the basis of your AI systems. Integration between different AI systems must be deeply understood, and the bias and limitations of the dependent AI implementations must be considered and driven by well-understood underlying data models. Establish feedback controls and mechanisms to analyze outcomes from the AI system to validate that all policies, security checks, attestations, quality, verifications, and system protection requirements have been enforced.

	Current	Target	Action
TECHNOLOGY	Today's use of AI (generally) leverages and reinforces existing knowledge but rarely creates significant new knowledge. It is reiterating and summarizing existing perspectives and solutions to problems.	AI gives meaning and understanding to difficult problems that are hard to model or understand as a human, and provides relevance to the outcome. AI acts as a layer to translate from human objectives to specialized outputs such as Python source code.	AI should be considered a complementary layer to existing data systems and should be subject to the same standards, contexts, governance, and operational controls of the organization.
	Validation of AI inferences and outcomes/decisions is immature. It is typically only performed at the AI System level and not at the business level.	Auditing of each step of the process is possible, and transparency enables validation for both business and consumers (either quantitatively or, more usually in AI, qualitatively). Established processes are used to run simulations on past data, capture narratives as feedback to ensure something makes sense qualitatively, learn from the past (ensure that the AI model is learning) and improve accuracy / reliability).	Define failsafe gateways on outcomes (fail-open or fail-closed) based on transaction traceability (what was used initially, then later, how it changes over time, and why). This technology must support the incident process for dealing with outcomes that vary outside of defined guardrails. Ensure that outcomes in the AI system can be readily understood and explained and/or decisions made by AI can be reversed. Establish controls/mechanisms to analyze the integrity of data models for actions against bias and limitations. The AI implementation must be deployed with measurability and observability tooling to enable reporting to support the governance requirements. This reporting includes enabling traceability and measurement of training data and outcomes.

USEFUL SOURCES:

- Biden-Harris Administration Executive Order Directs DHS to Lead the Responsible Development of Artificial Intelligence https://www.dhs.gov/news/2023/10/30/fact-sheet-biden-harris-administration-executive-order-directs-dhs-lead-responsible
- NIST Technical AI Standards https://www.nist.gov/artificial-intelligence/technical-ai-standards#:~:text=A%20broad%20spectrum%20of%20standards,for%20trustworthy%20and%20responsible%20AI.
- European Union AI Act: https://artificialintelligenceact.eu/the-act/
- UK AI Safety Summit topics and recommendations: https://www.gov.uk/government/topical-events/ai-safety-summit-2023

METHODOLOGY

To craft this opinion piece, we devised a structured process for distilling key aspects and considerations from the vast landscape of turbulence in the realm of Artificial Intelligence development, adoption, and usage while also incorporating AI within the process itself.

We initiated the process by collecting information from various sources, including research findings, emerging legislative frameworks, insights drawn from our own experiences, and contributions from industry experts. The collected information formed the foundation of our dataset. Subsequently, we harnessed the Azure Open AI Studio service [gpt-3.5-turbo-16k, version 0613] with the grounded dataset, utilizing it to generate initial draft versions of the paper.

The output generated by the AI was meticulously examined, and revisions were introduced wherever we found discrepancies or concerns with the AI-generated content. Following this internal review, the document was disseminated for scrutiny and commentary across various industry stakeholders. After receiving invaluable feedback, we engaged in a series of in-depth reviews and applicability assessments, ultimately culminating in the creation of the final version of this positioning paper.