The Challenges of Artificial Intelligence in Organizational Leadership and the Importance of Human-In-The-Loop Authority

Anthony R. Maranto, Ph.D.

Akamai University, Environmental Science Program, Kamuela, Hawaii and Durham, North Carolina.
Shenandoah University, Transformative Leadership Program, Leesburg, Virginia.

E-mail: dr.anthony.maranto@outlook.com

ABSTRACT

There is a large and growing body of literature examining how artificial intelligence (AI) will change the workforce of tomorrow and how it will impact organizational leadership in that context. This paper seeks to highlight some of the limitations of AI as it is currently being developed and implemented and how those limitations may influence the degree to which it may be employed as an augmentative tool or, conversely, what complications will need to be addressed before it can be truly seen from a replacement perspective. This paper presents a concise summary from the literature of some of the most fundamental leadership challenges to large scale implementation of AI in complex organizations.

(Keywords: transformative leadership, organizational leadership, artificial intelligence, machine learning, AI/ML, organizational change)

INTRODUCTION

The rapid increase in the networking of complex IT systems has resulted in enormous amounts of data being generated on a daily basis. In 2023, it was estimated that the world created around 120 zettabytes (ZB) of new data which roughly translates to 337,080 petabytes (PB) of daily data to support the needs of 5.35 billion global internet users (Edge Delta, 2024).

To take advantage of this massive amount of data, research institutions, governments, and industry have turned increasingly to artificial intelligence (AI) techniques. AI has been viewed as a truly disruptive technology with major implications for the workforce and organizational leaders. AI combines aspects of both engineering and cognitive science and depending upon the application, different scientific disciplines such as speech recognition, natural language processing, and computer vision can be integrated (Peifer, Jeske, and Hille, 2022). It is estimated that by 2030, AI systems will contribute up to $15.7 trillion to the global economy (De Cremer and Kasparov, 2021).

A distinction is often made between Artificial Narrow Intelligence, Artificial General Intelligence and Artificial Super Intelligence as general thresholds in the AI spectrum (Figure 1). Most current industrial applications are specific in nature and only operate within the boundaries of Artificial Narrow Intelligence and cannot be said to exceed normalized human intelligence (Peifer, Jeske, and Hille, 2022). As applications and systems develop and integrate, however, AI as a whole is rapidly advancing to the point of Artificial Super Intelligence.

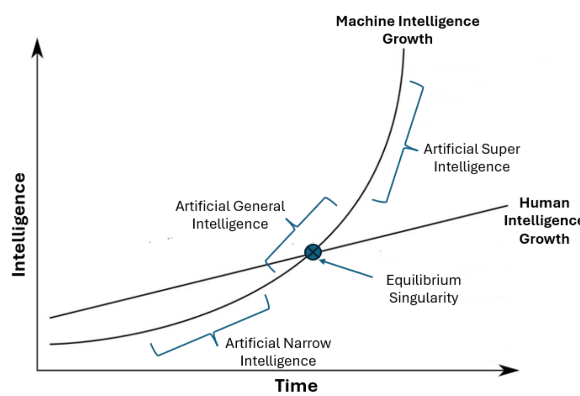

Figure 1: Phases of Growth for Machine Intelligence (adapted from Liebert and Talg, 2018).

Neither machine learning (ML), machine intelligence, nor the challenges of societal adaptation to them are new concepts. Since

Charles Babbage introduced his analytical engine in 1837, the computational power of mechanical devices has been well known, and futurists and philosophers have contemplated the possibility of intelligent, self-directed thought within machines long before it became a reality.

As early as 1950, Alan Turning of the UK National Physical Laboratory presented a detailed examination of the concept of machine intelligence and philosophically dismissed the question of independent machine intelligence, concluding that machines were only capable of the imitation of human thought (Turning, 1950). Innovations in machine learning and artificial intelligence, however, have advanced rapidly beyond the point of Turning's initial examination of the topic (DeCanio, 2016).

Effective applications of narrow AI have grown extensively in recent years. Once limited to highly mathematical predictive endeavors such as multiple regression applications for risk-based prioritization or cost prediction, new applications are quickly expanding to include the development of virtual assistants; image and speech recognition; Natural Language Processing (NLP); autonomous vehicle controls; and medical diagnostics (IBM, 2024; Galarraga, Werle, and Maranto, 2001; Dzuary and Maranto, 1999).

As with any complex human endeavor, the direction and leadership of diverse organizations marry an alignment of management, vision, economics, politics, and policy (Maranto, 2024; Maranto, 2019). Organizations can choose to not embrace trends in technologies, but often suffer in the long run when they select this path either through conscious decision making or through inaction.

In the 1980s, academics began fully examining the possibilities that some forms of artificial intelligence could conceivably be more successful than humans in motivating organizations and staff, minimizing the disturbances in information flow, dealing with organizational budgets, and addressing organizational mission objectives (Willmer, 1986; Alexander, 1984).

Ali Naqvi of the American Institute of Artificial Intelligence provided an assessment of how AI might influence the workplace of the near future and stated that AI will "create never-seen-before challenges for leadership". Naqvi's research highlighted two phases of leadership which will require directed analysis. The first was related to leadership focused on the great transformational changes required to transition from the industrial to the cognitive economy. The second area was related to the leadership required once firms reached a relative degree of stability and maturity where intelligent machines occupy a central position within the workforce (Naqvi, 2017).

In 2018, Zhaohao Sun of the Papua New Guinea University of Technology examined the nature of management and leadership roles and proposed an artificial leadership multi agent system that could preform the most fundamental aspects of leadership, namely changing strategically, influencing others, assigning responsibility, and delegating authority (Sun, 2018).

Van Quaquebeke of the University of Exeter and Gerpott of the Otto Beisheim School of Management, speculate that AI will not only support but come to substitute human leadership, "completely assuming authority over task-, relations-, and change-oriented functions that people prototypically associate with human leaders". Their research suggests that AI could be more efficient at fulfilling employees' needs for autonomy; fulfilling employees' needs for competence; and fulfilling employees' needs for relatedness. They suggest that "very likely, we will need fewer human leaders, particularly at the lower and middle management levels because their leadership functions can easily be taken over by AI and maybe for the better because we have seen that fewer people want or are even willing to take on leadership roles" (Van Quaquebeke and Gerpott, 2023).

The Center for Creative Leadership conducted a comprehensive study in 2020, during the first year of a disruptive event (the global COVID-19 pandemic), to assess how organizational leaders were coping with disruptions to their organizational missions. The results of that survey yielded five major disruption categories that were common across the hundreds of respondents and various business sectors, namely, big data and analytics; developing an innovative culture; changes associated with artificial intelligence; equity, diversity, and inclusion; and communications overload (Beckert and Jones, 2021).

There is a large and growing body of literature examining how AI will change the workforce of tomorrow and how it will impact organizational leadership in that context. Much of the literature can best be described as falling into one of three major perspectives: 1) a skeptical perspective indicating that the technology will never replace human thought, leadership, or creativity; 2) a replacement perspective that concludes that AI and robotic workforces will replace both leaders and followers in the modern economy; and 3) an enhancement perspective, where AI is viewed as an augmentative tool to leadership functions (Titareva, 2021).

This third view of the augmentative nature of AI technology was summarized nicely in the *Harvard Business Review* by David De Cremer, Provost at the National University of Singapore and Gary Kasparov, the chess Grandmaster who's famous matches against the IBM super-computer, Deep Blue, in the 1990s were instrumental in bringing the advances of AI to the mainstream. De Cremer and Kasparov (2021) in an examination of whether humans and machines are really in competition with each other stated, "the question of whether AI will replace human workers assumes that AI and humans have the same qualities and abilities — but, in reality, they don't. AI-based machines are fast, more accurate, and consistently rational, but they aren't intuitive, emotional, or culturally sensitive. And, it's exactly these abilities that humans possess and which make us effective."

This paper seeks to highlight some of the limitations of AI as it is currently being developed and implemented that may influence the degree to which AI may be employed as an augmentative tool or, conversely, what complications will need to be addressed before it can be truly seen from a replacement perspective. Presented below is a concise summary from the literature of some of the most fundamental leadership challenges to large scale implementation of AI in complex organizations.

METHODOLOGY

This paper attempts to condense from the research literature, challenges that have been expressed by researchers and organizations seeking to implement AI to support or enhance the leadership of large and complex organizations. In this research, the author relies upon content analysis techniques and cited evidence from data sources that have reported on the application of AI in a variety of organizational environments, data systems, and business settings.

Source material was identified using search parameters in the literature associated with artificial intelligence, machine learning, business leadership, organizational leadership, transformative leadership, and disruptive organizational change. A large portion of the current literature involving the application of AI originates from commercial entities with scholarly research being focused on technical developments in this field. The author reviewed 80 full manuscripts and over 200 abstracts and non-refereed case use reports in the compilation of this paper.

RESULTS AND DISCUSSION

Algorithmic Transparency and Bias

One of the first challenges that relates to the implementation of AI for leadership of complex organizations relates to the need to ensure transparency in AI algorithms. Complex models may lack interpretability, making it difficult to understand and dissect their decision-making process.

It is expected that to be able to fully take advantage of AI systems as a decision-making tool, leaders must have at least a basic understanding of the functioning of AI as these systems tend to be interconnected and complicated to understand. In order to implement AI successfully in organizational leadership, decision makers need a comprehensive understanding of the AI system operations and functions to avoid the possibility of compounding biases from incomplete or skewed learning scenarios or feeder systems (Björkman and Johansson, 2018). AI systems may be able to rapidly come to conclusions based on information compiled from massive datasets, but if the humans-in-the-loop are not able to articulate how those decisions were made and what assumptions accompany them, confidence in those decisions will be undermined and the ultimate data products and decision support tools used may misrepresent outcomes (Brynjolfsson and McAfee, 2017).

Mitigating bias in AI algorithms is challenging and requires proactive measures to ensure fairness and equity in decision-making processes. AI, ML, and algorithmic bias are terms that refer to AI systems that produce skewed or false results that reflect the biases of their underlying databases or foundational systems. These may include long standing logical errors or may include biases that are based on social constructs. These biases may be present in the training data used for ML, the AI algorithms, or within the predictions that those algorithms produce (IBM, 2023; Chouldechova, 2017).

When these biases go uncorrected, the decision systems may produce distorted results which undermine the confidence in future decisions. Correction requires forensic examination of training datasets and AI/ML algorithm logic.

The IBM Artificial Intelligence Team summarizes the basic classifications of AI bias as (IBM, 2023):

Training Data Bias: AI systems learn their pattern recognition based on large training datasets. It is therefore essential to assess datasets for the presence of bias, skewedness, or conditional limitations. Data sampling should be utilized to identify over- or under-represented groups within the training datasets that may come from biased population and demographic data or geographical over-representation. Training bias can also result from how the training data is labeled (i.e., inconsistent labeling of data characteristics or search parameters).

Algorithmic Bias: Flawed algorithms can result in consistent (or inconsistent) errors in data assimilation or model output, biased predictive outcomes, and may even compound the bias inherent in a flawed dataset. It may be the result of programming errors, faulty logic, or over/under-weighting of evaluation factors or risk elements in the decision-making system.

Cognitive Bias: As end-users of decision support tools or data products make choices and judgements based on model output, they are inevitably influenced by their experiences and preferences. As a result, people may build these biases into AI systems through the selection of data or how the data is weighted. For example, cognitive bias could lead to favoring datasets gathered from Americans rather than sampling from a range of populations around the globe.

The US Department of Commerce's National Institute of Standards and Technology (NIST) released a technical report titled, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence", that noted that "human and systemic institutional and societal factors are significant sources of AI bias as well and are currently overlooked". The report stated that "successfully meeting this challenge will require taking all forms of bias into account. This means expanding our perspective beyond the machine learning pipeline to recognize and investigate how this technology is both created within and impacts our society." (Schwartz, *et al.*, 2022)

There have been many examples of lack of transparency in AI algorithms and bias within their construction in the literature. In 2019, Obermeyer, *et al*., reported in *Science* on evidence of racial bias in a widely used health care algorithm, which showed that African American patients assigned the same level of risk by the algorithm were sicker than Caucasian patients. That bias occurred because the algorithm used health costs as a proxy for health needs, and less money was spent on African American patients who had the same level of need (Obermeyer, *et al*., 2019).

In another study in *Nature, Communications Medicine*, Adam, *et al*., showed how AI system models can exhibit problematic biases for emergency services. Their study used a web-based experiment to evaluate the impact biased models can have when used to inform human decisions. In this experiment, adherence to AI recommendations led evaluators to be biased against African-American and Muslim individuals (more frequently recommending law enforcement intervention) in responding to mental health crisis events (Adam, *et al*., 2022).

Even attempts to correct for perceived bias in datasets can have second order effects in AI outcomes. Google's Gemini AI chatbot received a great deal of scrutiny recently when AI generated graphics depicting a range of historically inaccurate images of Asian Nazis, African-American founding fathers, and female Popes became the subject of world-wide memes pushing back on the perceived cultural movement to rewrite history in an attempt to irradicate all historical traces of racial prejudice and discrimination (Figure 2). As a result of the backlash, Google CEO Sundar Pichai, responded in a memo to staff calling the photos 'problematic' and 'unacceptable' and instructing technical staff to work around the clock to fix the issues (Main, 2024).

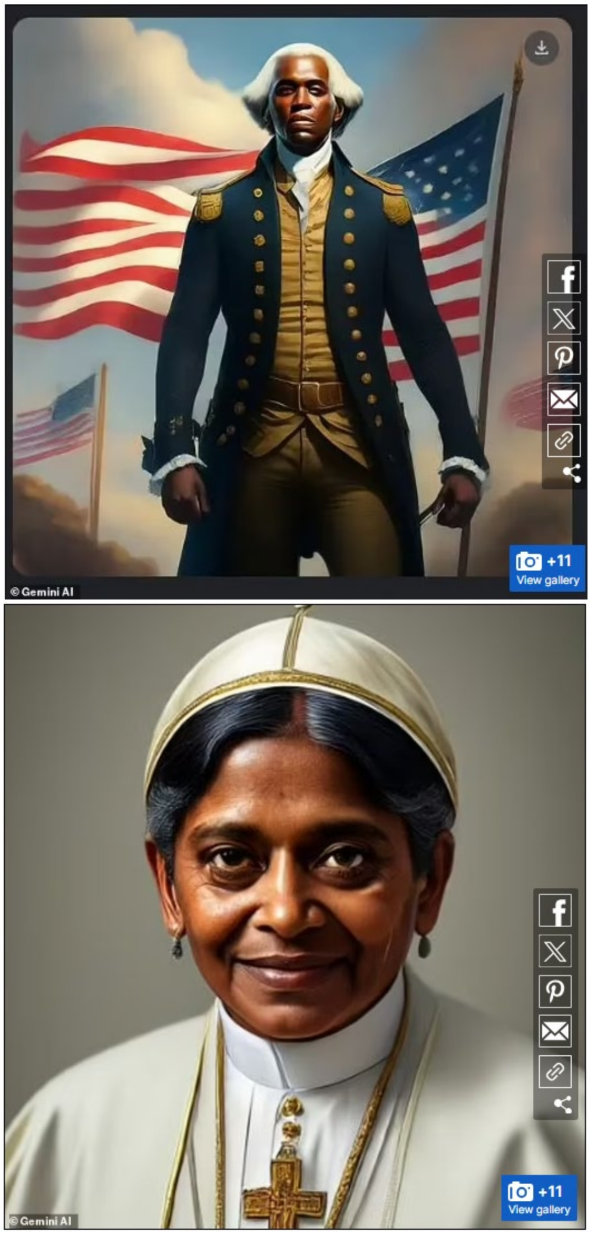

Figure 2: Historically Inaccurate Images Generated by Gemini AI of an African-American US Founding Father and a Female Catholic Pope. (from Main, 2024/*Daily Mail*)

Product Quality and Cybersecurity Threats

Without high-quality data, even the best ML algorithms will underperform expectations or provide questionable results and data products. In modern data-centric AI practices that reside in the Artificial Narrow Intelligence range, many real-world datasets have been characterized as small, dirty, biased, and even poisoned (Whang, *et al*., 2023).

Data collection and data quality becomes important in AI systems so that there is less need for future software/systems engineering for implemented deep learning approaches. A study from researchers from the Korean Advanced Institute of Science and Technology presented recommendations for data validation, cleaning, and integration techniques to help cope with imperfect data during model training for traditional data management research. They also recommend fairness measures and unfairness mitigation techniques that can be applied to datasets before, during, or after model training to tune model quality (Whang, *et al*., 2023).

With regards to data quality, the accuracy, completeness, timeliness, integrity, consistency, and relevance of the data used for training and testing AI models directly affects the performance and accuracy of those systems. Sub-standard data can lead to inaccurate or irrelevant outputs and products and impact end-use decision-making and associated organizational processes that are dependent upon those AI outputs. Ensuring high quality data is imperative for the reliability of AI decision support systems.

Aldoseri, *et al*. (2023) of the College of Engineering, Qatar University, recommend procedures that include data cleansing, validation, enrichment, and management to ensure that AI applications have the high-quality relevant, representative, and reliable data needed to produce high-fidelity outcomes. Additionally, AI systems require frequent and ongoing validation, verification, and output monitoring to ensure that data quality is consistent over time (Aldoseri, *et al*., 2023).

Data quality and anomaly resolution is a multidimensional discipline of information science and systems engineering that incorporates factors such as data accuracy, data completeness, data uniformity, timeliness, data relevance, and data uniqueness. Addressing data quality issues involves implementing robust anomaly detection mechanisms, setting appropriate thresholds for anomaly detection, root cause analysis, implementing automated remediation processes, and implementing appropriate human oversight that includes continuous monitoring and establishing appropriate feedback loops. These approaches can help resolve anomalies in real-time or near-real-time, reducing their potential impact on AI model outputs and operations (Batista, and Monard, 2018).

Benedikt, *et al*. (2020) advocated for automation as a means to make efficiency savings and speed up processing time in business activities, but also strongly stated that “there are situations where we need human interventions to maintain data quality. Therefore, we propose a human-in-the-loop solution and adopt a human-centered approach to design a user interface that allows human-machine interaction to be closely intertwined at every step of the process”.

In addition to data quality issues, AI systems can be vulnerable to cyberattacks, requiring organizational leaders to invest in robust cybersecurity measures to protect against a host of potential electronic threats. AI systems also rely on vast amounts of data which may contain Personally Identifiable Information (PII), economically sensitive data, protected intellectual property, medically confidential information, or even governmentally classified information. Protecting this data from unauthorized access, misuse, or breaches is essential for maintaining organizational and public trust as well as regulatory compliance. Organizational leaders must therefore implement robust data privacy and IT security measures to safeguard against potential risks associated with AI deployment.

AI is a double-edged sword that can both be used to strengthen cyber security and can likewise, be turned against it. On the protective side, AI applications have been used to improve cyber threat detection; develop predictive models for the identification of Phishing and malicious bot attacks; strengthen access controls; and detect anomalous network traffic, unusual code, and breach points (Malwarebytes, 2024). However, AI can also be used for the optimization of cyber-attacks; the generation of automated malware; theft of intellectual property and other AI models through network attacks; and data manipulation and data poisoning (Malwarebytes, 2024).

Additionally, generative AI can also produce unique materials that are very close imitations of actual events, artifacts, and personal communications. The New York Institute of Technology (2023) states that these “deepfakes” are a type of synthetic media that use artificial techniques such as face swapping and voice cloning, coupled with deep learning algorithms, to create highly realistic and deceptive videos, audio recordings, or images. They state that “these AI-generated media typically involve the manipulation of existing content, such as replacing a person's face in a video with someone else's or altering their speech in an audio recording” (NYIT, 2023). Deepfakes have become harder and harder to distinguish from real accounts without complex forensics and have increasingly been used in different types of crimes, financial fraud, intellectual property theft, and even national security threats.

A recent example of this was seen at Pikesville High School in suburban Baltimore, Maryland where a school athletic director was arrested and charged with using an AI deepfake recording to impersonate the school’s Principal on an audio clip that included racially insensitive and derogatory comments about students and staff, in an apparent attempt to get that Principal fired (Nguyen, 2024).

Given the range of sensitive, confidential, and protected information sources that can be accessed and manipulated by AI, careful consideration must be given to protecting these systems against malicious AI vulnerabilities, attacks, and unauthorized access (Dixit, Quaglietta, and Gaulton, 2021).

Ultimately, manipulation of data need only be minimal to have a significant impact on AI decision support outcomes. In a controlled environment, researchers at the Faculty of Information Science and Electrical Engineering, Kyushu University, demonstrated that modifying a single pixel on facial images ingested into an AI security recognition system significantly altered what the algorithm believed it was seeing (Su, Vargas, and Sakurai, 2019). This example emphasizes the importance of strong IT security strategies to protect data integrity.

Human-in-the-loop leadership remains important to ensure audits for any implemented AI systems, controlling PII and sensitive data shared through automation, ensuring robust organization data security plans, optimizing software and cyber security programs, and developing comprehensive vulnerability management and incident response programs that support the entire organization.

Change Management

Implementing AI initiatives often requires leaders to proactively initiate significant changes to organizational culture, vision, processes, and workflows, necessitating effective change management strategies. Embracing any AI-driven innovation may require a significant cultural shift within organizations, with leaders needing to foster a mindset of experimentation, innovation, risk-taking, and institutional learning.

Resistance to change, cultural inertia, fear of job loss, and general employee skepticism can pose significant barriers to successful AI implementation within any organization. Leaders must invest in change management strategies, stakeholder engagement, and training programs to foster a culture that embraces AI-driven innovation.

Implementing AI-driven change management across an organization involves negotiating a complex mix of issues which range from technical implementation and integration challenges to organizational hierarchy and cultural adaptations.

On the technical side, the integration of AI systems with existing IT, business process, and knowledge management systems can require overcoming incompatibilities, revamping existing infrastructure, and reallocation of technical assets. Organizational challenges can be equally daunting and involve internal resistance (from management and non-management personnel alike) stemming from concerns over job displacement or the incorporation of complex new technologies. The leadership challenges in overcoming this resistance can involve implementing change management, training programs, and mentorship (Kuriakose, 2023).

Employee skill gaps also can represent a significant barrier in many organizations as the implementation of AI in change management demands specific expertise in AI/ML, data analytics, data informatics, data assimilation, and system engineering. This may necessitate either significant re-training of existing staff or recruitment of new employee expertise, both of which can strain an organization's budgets and the resources available to execute the organization's core missions.

A study by Scott Anderson of Prosci looked at industry responses from October 2023 to determine reasons for organizational resistance to AI. A total of 84% of change practitioners surveyed stated that they were moderately familiar to very familiar with AI and AI applications. However, only 48% said they currently use it in their change management work. The top three responses relate most directly to an overall lack of understanding, inadequate experience with AI, and fear of unidentified risks (Figure 3) (Anderson, 2024).

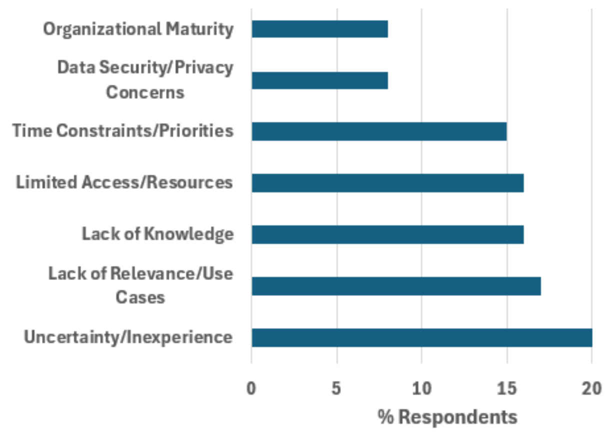

Figure 3: Primary Reasons for Not Using AI Technologies. (Source: Proseci Research Hub, adapted from Anderson, 2024)

In a recent study at Harvard University, Department of Computer Engineering, Gyanamurthy and Radhanath (2023) presented findings that showed the pivotal role of transformational leadership in driving AI adoption and innovation. The authors showed, "significant positive correlation between transformational leadership and AI adoption levels, indicating that leaders who exhibit visionary qualities and empower their employees are more likely to embrace AI technologies. This aligns with existing literature emphasizing the importance of transformational leadership in promoting organizational change and adaptability" (Gyanamurthy and Radhanath, 2023).

An industry whitepaper produced by the international consulting firm, Booz Allen Hamilton, distilled what they described as the five foundational steps needed to prepare an organization for the move towards innovative AI systems and solutions. The authors Katie Wrenn and Earnest Sohn indicated that organizational change should focus on the following areas in order to implement AI solutions geared towards improved performance, streamlined processes, and enhanced decision-making capabilities (Wrenn and Sohn, 2024):

Support Ambitious Goals with Small Steps: Establish a clear vision to show the connection between any AI success and the organization's goals. Set a vision and specific business objective for AI capability. Roll out AI capabilities incrementally, one project at a time, while staying focused on the mission. Encourage organizational "early adopters" to help prepare for change.

Human-Centered Design: Incorporate end users into all phases of the design and development lifecycle. Start with empathy and understanding for the needs of the end user. Gain insights through interviews, focus groups, and surveys. Include both technical end users and non-technical decision makers who will use the output.

Training and Knowledge Sharing: Tailor the change management plan to educate and train in a way that addresses specific adoption challenges. Conduct a stakeholder analysis to gain insights about staff reluctance. Anxiety about the systems, outcomes, or uncertain futures will cause employees to resist AI. Proactively address concerns in your AI Strategy.

Use Targeted Communications: Employ communications that support the same AI outcomes in tandem with change progress. Build a communications strategy around adopting AI which builds excitement while managing expectations. Maintain engagement and communications over the long haul. As AI becomes increasingly embedded in the organization, solicit employee feedback and lessons-learned.

Encourage Staff-to-Staff Dialogue: Create a sense of shared responsibility and support to incorporate AI. Create a cross-functional, co-educational, and interactive environment focused on the successful integration of AI tools. Utilize online forums or listserv as well as interactive opportunities, such as hack-a-thons, hands-on training sessions, or pilot showcases.

Interdisciplinary Collaboration

Successful AI initiatives often require collaboration across a wide range of disciplines, creating challenges in fostering communication and understanding between technical and non-technical teams. In the implementation and integration of an organizational AI strategy, the team needs to focus on the merger of data science, business analytics, upskilling the existing workforce, sharing cross-organizational data and insights, employee training and development, and developing new AI workflows and automation processes. As a result, organizations increasingly must equip themselves with a wide array of skills to effectively harness the potential of these new systems (Franz, 2024).

Patrice Maniglier of Paris Nanterre University offered an examination of the challenges of developing new interdisciplinary fields, such as with AI technology ethics, where intellectual encounters have not yet been fully developed to support the clearinghouse of shared terminology, definitions, and concepts needed to support effective collaboration (Moats, Holtrop, and van Eck, 2024; Maniglier, 2021)

Foundational questions around the interdisciplinary interactions of AI involve diverse topics including the assessment of valid methodologies, architectural concepts, and evaluative tools which are sometimes shared between academic disciplines but are often divergent with respect to the definitions and understandings of the required solutions (Moats, Holtrop, and van Eck, 2024; Coeckelbergh, 2020).

Interdisciplinarity is also required in the refinement of AI systems, algorithms, and data outputs/products. In a Harvard Business Review article, H.J. Wilson and Paul Daugherty of Accenture describe the multi-disciplinary aspects of human involvement required to develop and refine AI systems. Namely, they must train machines to perform the variety of tasks being assigned; they must explain the outcomes of those tasks (especially when results are controversial); and they must sustain the responsible use of machines (ensuring the containment of negative or unintended effects) (Wilson and Daugherty, 2018).

Their research examining 1,075 companies across 12 industries showed that organizations benefit from optimizing collaboration between humans and artificial intelligence and implementing an interdisciplinary development and refinement approach. Interdisciplinary principles of this approach focused on reimagining business processes; embracing experimentation; encouraging employee involvement; active planning for an AI strategy; responsibly collecting data; and redesigning workflows to incorporate AI (Wilson and Daugherty, 2018). These factors which combine technical, managerial, human resources, and organizational engineering all work together as needed elements in the development and integration process.

Ethical Dilemmas

Modern AI applications can raise complex ethical questions, particularly when used in leadership roles where decisions can have significant social, economic, and environmental impacts. Leaders must consider the ethical implications of AI deployment, including issues such as privacy, consent, accountability, and the potential for unintended consequences. Balancing the desire for efficiency and productivity with ethical principles can be challenging, especially in environments where short-term gains may overshadow long-term risks.

A large part of the advancements in AI have been made in the field of ML thanks to the availability of Big Data. It has been suggested, therefore, that AI ethics are truly seen as a sub-component of the field of data ethics (Floridi and Taddeo, 2016). There have been similar ethical discussions around the development of Big Data when the ethical examination moved beyond PII data and towards considerations of how personal and non-personal data were used especially in commercial applications, political analytics, and research applications. As systems integrated and developed increasing degrees of autonomy the AI ethics discussion has drawn in other aspects of applied ethics as well as moral theory, political philosophy, and bioethics (Moats, Holtrop, and van Eck. 2024; Bietti, 2020; Mittelstadt, 2019, Binns, 2018).

Applications in specific fields may also hold specific and complex ethical considerations. For example, AI applications in the healthcare and biomedical sectors must be compliant with a long list of state and country-specific regulatory and ethical guidelines which have evolved on a rapid and frequent basis. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) has strict requirements for the protection and releasability of medical data. These regulations are mirrored in many countries by such legislation as the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, the Privacy Act 1988 in Australia, and the General Data Protection Regulation (GDPR) throughout the European Union.

Efforts to adopt AI in these fields also must address inequities that may exacerbate gaps in the quality of or access to medical services or clinical trials since biases in those datasets or decision support tools may represent a risk to the ethical adoption of AI analytics. Dixt, Quaglietta, and Gaulton (2021) indicate that a critical step to ensuring ethical application of AI solutions is to ensure the use of appropriate and representative data and avoid or apply corrections for historical bias, representation bias, measurement bias, aggregation bias, evaluation bias, and development bias (Yu and Kohane, 2019; Dixit, Quaglietta, and Gaulton, 2021)

The use of AI in leadership functions may be subject to additional regulatory requirements aimed at ensuring fairness, transparency, and accountability of supporting services and tools. Compliance with these regulations adds an additional layer of complexity to AI implementation, as organizations must navigate a patchwork of legal frameworks that may vary across jurisdictions and industries.

Within the military context, where autonomous decisions by AI systems may have significant and life-threatening repercussions on a geo-political scale, the ethical considerations of AI systems become very apparent. In December 2015, the US Deputy Secretary of Defense, Robert Work, said in a symposium held by the Center for New American Studies, "…we believe, strongly, that humans should be the only ones to decide when to use lethal force. But when you're under attack, especially at machine speeds, we want to have a machine that can protect us" (Beckert and Jones, 2021).

In follow up discussions and examinations resulting from Work's presentation, the topic became euphemistically known as the "human-in-the-loop" speech. The Department of Defense firmly established that although emerging technologies including AI/ML could identify threats, assess scenarios, and launch reactions faster than any human decision-making chain, the Department would not allow fully autonomous attack or counter-attack scenarios. Human leaders applying well established levels of control and responsibility, chain-of-command, and human ethics, judgment, and national values, would ultimately make use-of-force decisions (Beckert and Jones, 2021).

Within the business world, similar challenges (although without the use of lethal force element) exist every day. Lessons learned from business experience, when augmented by rapid machine pattern detection and processing, should be addressed with the application of strategic insights and ethical direction to be supportive as business leaders make critical decisions (Beckert and Jones, 2021). This may ultimately require a human-in-the-loop element and an intentional slowing of the speed of decision making.

CONCLUSION AND RECOMMENDATIONS

AI development clearly has the potential to apply advanced analysis and logic-based computational techniques which may be used to solve complex societal problems, with human-like intelligence and independence (Alsheibani, *et al*., 2020). However, AI technology is immature, with known challenges in the areas of algorithmic transparency, product quality and security, change management, interdisciplinary collaboration, and in the area of addressing ethical dilemmas.

Addressing these challenges requires a holistic approach that combines technological expertise, ethical considerations, and creative organizational science. By proactively addressing transparency in AI algorithms and algorithm bias, leaders can harness the transformative potential of AI while minimizing risks and maximizing benefits for their organizations and stakeholders.

Given that the mass of AI technology is rapidly advancing towards Artificial General Intelligence levels and can be expected to match or exceed human cognition across a wide range of tasks, a critical concern requires the degree of autonomy that should be targeted in AI architecture.

AI systems are fundamentally faster in computational processing than human-based systems and they also have benefits in accuracy, efficiency, and reproducibility. They do, however, have limitations to their ability to process information contextually and can often fail in stunning ways when extrapolating scenarios and novel situations outside of their training sets. Allowing these systems to operate within the guard-rails of human oversight (i.e., human-in-the-loop) provides a mechanism to supervise AI systems and outputs on impactful or high-risk actions (Mazzolin, 2020). Requiring an AI system output to attain human approval prior to the execution of the system's chosen course of action, allows for a final intervention step should something go wrong, or if a more detailed dissection of the recommendation is required.

Until a combination of industry best practices and regulatory guidelines has been established to provide the guiderails for truly autonomous AI systems and decision workflows, a human-in-the-loop aspect of authority and decision approval should be considered for the most important decision processes, especially those that involve significant financial, resource, or health and safety consequences. It is therefore anticipated that AI systems will continue to support organizations in an enhancement rather than replacement mode for the foreseeable future.

Recommendations for future research and implementation include:

- Working with both academia and industry to identify and address AI bias through governance or the improved ability monitor large AI datasets or products.

- Encouraging AI-related development to proceed with consistent and relevant industry regulations, legal parameters, and ethical guidelines.

- Fostering industry-wide cooperation to protect customers' information and create trustworthy and secure AI systems.

- Ensuring algorithm transparency to avoid the fielding of complex "black box" decision systems which offer little insight into the data used, the weighting assigned, or the logical output.

- Consistent consideration of safeguards like the "human-in-the-loop" systems to make recommendations that are then reviewed by humans before an action is initiated.

- Future changes in leadership should be redefined to not establish an endpoint of AI replacing human leadership, but rather, on how leadership will evolve to function as an AI monitor for better decision-making and how leaders of the future will serve as facilitators of AI systems for collaboration and creative solution architecture.

REFERENCES

1. Aldoseri, A., K.N. Al-Khalifa, and A.M. Hamouda. 2023. "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges". *Applied Science*. 13(12): 7082. https://doi.org/10.3390/app13127082

2. Adam, H., A. Balagopalan, E. Alsentzer, F. Christia, and M. Ghassemi. 2022. "Mitigating the Impact of Biased Artificial Intelligence in Emergency Decision-Making". *Nature: Communications Medicine*. 2: 149. https://doi.org/10.1038/s43856-022-00214-4

3. Alexander, T. 1984, "Why Computers Can't Outthink the Experts". *Fortune*. 110(4).

4. Alsheibani, S.A., C. Messom, Y. Cheung, and M. Alhosni. 2020. "Reimagining the Strategic Management of Artificial Intelligence: Five Recommendations for Business Leaders". *AMCIS 2020 Proceedings*. 4. https://aisel.aisnet.org/amcis2020/is_leadership/is_leadership/4

5. Anderson, S. 2024. "AI in Change Management: Early Findings, Challenges and Opportunities". Proseci: Fort Collins, CO. https://www.prosci.com/blog/ai-in-change-management-early-findings

6. Batista, G.E. and M.C. Monard. 2018. "Data Quality in Machine Learning: A Study in the Context of Imbalanced Data". *Neurocomputing*. 275: 1665–1679.

7. Beckert, C. and P. Jones. 2021. "Human in the Loop: Leadership in a Disrupted Environment". *The Journal of Character and Leadership*. (Summer): 16-21.

8. Benedikt, L., C. Joshi, L. Nolan, R. Henstra-Hill, L. Shaw, and S. Hook. 2020. "Human-in-the-Loop AI in Government: A Case Study". In: 25th International Conference on Intelligent User Interfaces (IUI). March 17–20, 2020. Cagliari, Italy. https://doi.org/10.1145/3377325.3377489

9. Bietti, E. 2020. "From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Barcelona, Spain. 210–219.

10. Binns, R. 2018. "Fairness in Machine Learning: Lessons from Political Philosophy". In: Conference on Fairness, Accountability and Transparency. PMLR: New York, NY. 149–159.

11. Björkman, I. and S. Johansson. 2018. "What Impact will Artificial Intelligence have on the Future Leadership Role? – A Study of Leaders' Expectations". Lund University, School of Economics and Management: Lund, Sweden.

12. Brynjolfsson, A. and J. McAfee. 2016. *The Second Machine Age; Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Norton & Company: New York, NY.

13. Chouldechova, A. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments". *Big Data*. 5(2): 153–163.

14. Coeckelbergh, M. 2020. *AI Ethics*. The MIT Press: Cambridge, MA. 248pp. ISBN: 9780262538190

15. DeCanio, S.J. 2016. "Robots and Humans: Complements or Substitutes?". *Journal of Macroeconomics*. 49: 280–291.

16. De Cremer, D. and G. Kasparov. 2021. "AI Should Augment Human Intelligence, Not Replace It?". *Harvard Business Review: Economics and Society*. March 18, 2021

17. Dixit, A., J. Quaglietta, and C. Gaulton. 2021. "Preparing for the Future: How Organizations can Prepare Boards, Leaders, and Risk Managers for Artificial Intelligence". *Healthcare Management Forum*. 34(6): 346–352.

18. Dzuray, E.J. and A.R. Maranto. 1999. "Assessing the Status of Risk-Based Approaches for the Prioritization of Federal Environmental Spending" *Federal Facilities Environmental Journal.* 10(2): 25 – 42. DOI: 10.1002/ffej.3330100204

19. Edge Delta. 2024. "Breaking Down the Numbers: How Much Data Does the World Create Daily in 2024?". Edge Delta: Seattle, WA. https://edgedelta.com/company/blog/how-much-data-is-created-per-day#:~:text=Breaking%20down%20the%20figures%20will,15.87%20TB%20of%20data%20daily.

20. European Commission. 2020. "AI Watch: Artificial Intelligence in Public Services". European Commission: Brussels, Belgium. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120399/

jrc120399_misuraca-ai-watch_public-services_30062020_def.pdf

21. Floridi, L. and M. Taddeo. 2016. “What is Data Ethics?”. *Philosophical Transactions of the Royal Society*. 374 (2083): 1-5. DOI:10.1098/rsta.2016.0360

22. Franz, A. 2024. “Skills Employees Need in order to Leverage Generative AI”. CX Journey. https://cx-journey.com/2024/06/skills-employees-need-in-order-to-leverage-generative-ai.html

23. Galarraga, H.E., C.P. Werle, and A.R. Maranto 2001. “Predicting Future Environmental Funding Needs: Regression Modeling of Historic Project and Program Costs”. *Federal Facilities Environmental Journal*. 12(3): 25-37. DOI: 10.1002/ffej.1014

24. Gyanamurthy, V. and S. Radhanath. 2023. “Leading Change in the AI Era: Strategies for Transformational Leadership”. *International Journal of Advanced Engineering Technologies and Innovations*. 1(4): 164-184.

25. Hagendorff, T. 2020. “The Ethics of AI Ethics: An Evaluation of Guidelines”. *Minds and Machines*. 30: 1–22.

26. IBM (International Business Machine). 2024. “What Is Narrow AI and What Are Its Applications?”. Big Blue Data Academy: Data Science Stories. February 11, 2024. https://bigblue.academy/en/narrow-ai#:~:text=Narrow%20AI%20is%20highly%20effective,sorting%20and%20categorizing%20different%20faces.

27. IBM (International Business Machine). 2023. “Shedding Light on AI Bias with Real World Examples”. *Think Blog*. October 16, 2023. https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/

28. Kuriakose, A.A. 2023. “AI-Driven Change Management in ITSM: Challenges and Solutions”. Algomox: Karnataka, India. https://www.algomox.com/resources/blog/ai-driven-change-management-in-itsm-challenges-solutions/

29. Liebert, K. and A. Talg. 2018. “Künstliche Intelligenz und das Lernen der Zukunft (Artificial Intelligence and the Learning of the Future)”. In: *HR-Trends 2019*. Strategie, Digitalisierung, Diversität, Demografie. K. Schwuchow and J. Gutmann (eds.). Haufe Group: Freiburg im Breisgau, Germany.

30. Maniglier, P. 2021. “Problem and Structure: Bachelard, Deleuze, and Transdisciplinarity”. *Theory, Culture and Society*. 38(2): 25–45.

31. Malwarebytes. 2024. “AI in Cyber Security: Risks of AI”. Malwarebytes Cyber Resource Center: Santa Clara, CA . https://www.malwarebytes.com/cybersecurity

32. Maranto, A.R. 2024. “The Boundary Layer between Science and Policy in Public Decision Making”. *Pacific Journal of Science and Technology*. 25(1): 120-131.

33. Maranto, A.R. 2019. “Where Science Meets Policy: The Nexus of Decision Making in Modern Society”. In: *Educational Transformation*. Akamai University: Hilo, HI. 10-28.

34. Maranto, A.R. 1998. “Classic Barriers to Novel Science”. *Exotic Research Report*. 2(4): 61-64.

35. Mazzolin, R. 2020. “Artificial Intelligence and Keeping Humans ‘in the Loop’”. Center for International Governance Innovation: Waterloo, Canada. November 23, 2020. https://www.cigionline.org/articles/artificial-intelligence-and-keeping-humans-loop/

36. Mittelstadt, B. 2019. “Principles Alone Cannot Guarantee Ethical AI”. *Natural Machine Intelligence*. 1: 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4

37. Moats, D., T. Holtrop, and N.J. van Eck. 2024. “Making Problems: Interdisciplinary Collaboration and AI Ethics”. Open Science Foundation: Charlottesville, VA. https://osf.io/hgdmt/download/?format=pdf

38. Naqvi, A. 2017. “Responding to the Will of the Machine: Leadership in the Age of Artificial Intelligence”. *Journal of Economics Bibliography*. 4(3): 244- 250.

39. Nguyen, T. 2024. “Baltimore High School Athletic Director used AI to Create Fake Racist Recording of Principal, Authorities Say”. *USA TODAY*. 4/24/24. https://www.usatoday.com/story/news/nation/2024/04/25/baltimore-maryland-artificial-intelligence-teacher-deep-fake/73460123007/

40. Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. “Dissecting Racial Bias in an Algorithm used to Manage the Health of Populations”. *Science.* 366(6464): 447-453. DOI: 10.1126/science.aax2342

41. Peifer, Y., T. Jeske, and S. Hille. 2022. “Artificial Intelligence and its Impact on Leaders and Leadership”. *Procedia Computer Science*. 200: 1024–1030.

42. Schwartz, R., A. Vassilev, K. Greene, L. Perine, A. Burt, and P. Hall. 2022. “Towards a Standard

for Identifying and Managing Bias in Artificial Intelligence". NIST Special Publication 1270. NIST: Washington, D.C. https://doi.org/10.6028/NIST.SP.1270

43. Su, J., D.V. Vargas, and K. Sakurai. 2019. "One Pixel Attack for Fooling Deep Neural Networks". *IEEE Transactions on Evolutionary Computation*. 23(5): 828-841. doi:10.1109/TEVC.2019.2890858.

44. Sun, Z. 2018. "Artificial Leadership". *PNG University of Technology BAIS*. 3(12): 1-7.

45. Titareva, T. 2021. "Leadership in an Artificial Intelligence Era". Presented at Leading Change Conference 2021. James Madison University: Harrisonburg, VA. February 5, 2021.

46. Turing, A.M. 1950. "Computing Machinery and Intelligence". *Mind.* 59(236): 433–460 .

47. Van Quaquebeke, N. and F.H. Gerpott. 2023. "The Now, New, and Next of Digital Leadership: How Artificial Intelligence (AI) Will Take Over and Change Leadership as We Know It". *Journal of Leadership and Organizational Studies*. 30(3): 265–275.

48. Whang, S.E., Y. Roh, H. Song, and J.G. Lee. 2023. "Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective". *The VLDB Journal: The International Journal on Very Large Data Bases*. 32: 791–813.

49. Willmer, M.A.P. 1986. "Can Artificial Intelligence Do Better than Humans at Leadership?". In: *Artificial Intelligence in Economics and Management*. L.F. Pau (Ed.). Elsevier Science Publishers: Amsterdam, The Netherlands.

50. Wilson, H.J. and P. Daugherty. 2018. "Collaborative Intelligence: Humans and AI Are Joining Forces". *Harvard Business Review*. July–August 2018: 114–123.

51. Wrenn, K. and E. Sohn. 2024. "Engineering your AI Future: Navigating AI Change Management like a Boss". Booz Allen Hamilton: McLean, VA. https://www.boozallen.com/insights/ai/change-management-for-artificial-intelligence-adoption.html

52. Yu, K.H. and I.S. Kohane. 2019. "Framing the Challenges of Artificial Intelligence in Medicine." *British Medical Journal (Quality & Safety in Health Care)*. 28(3): 238-241.

ABOUT THE AUTHOR

Dr. Anthony Maranto, has over 30 years of experience as a researcher, educator, program manager, and business leader. He has spent over 14 years as a consultant for the National Oceanic and Atmospheric Agency, National Aeronautics and Space Administration, and the National Institutes of Health. He has also spent 12 years as a consultant to the US Department of Defense where he was extensively involved in numerous technology, compliance, and environmental programs as well as numerous projects in the Chemical, Biological, Radiological, and Nuclear (CBRN) defense arena. Dr. Maranto has held academic and administrative posts at several universities and colleges. As a researcher, he has published or presented over 50 technical papers and reports on numerous topics including environmental management, strategic planning, carcinogenic risk assessment, radioecology, health physics, data quality control, risk based funding, and neurological receptor modeling.

SUGGESTED CITATION

Maranto, A.R. 2024. "The Challenges of Artificial Intelligence in Organizational Leadership and the Importance of Human-In-The-Loop Authority". *Pacific Journal of Science and Technology*. 25(2): 81-93.

Pacific Journal of Science and Technology