THE AI TRANSFORMATION RISK MODEL: SOLUTIONS WHERE BENEFITS OUTWEIGH RISKS

Richard Fulton,[1] Diane Fulton,[2] Nate Haynes[3] and Susan Kaplan[3]

[1] Department of Computer Science, Troy University, Troy, Alabama, USA
[2]Department of Management, Clayton State University, Morrow, Georgia, USA
[3]Modal Technology, Minneapolis, Minnesota, USA

ABSTRACT

This paper summarizes the most cogent and recently-cited advantages and risks associated with Artificial Intelligence from an in-depth review of the literature. Then the authors synthesize the salient risk-related models currently being used in AI, technology andbusiness-related scenarios. Lastly,in view of the most pressing issues and updated context of AI along with theories and models reviewed, the writers proposea new framework called "The AI Transformation Risk Model"to address the increasing fears and levels of risk. Using the model characteristics as a backdrop, the article emphasizes innovative solutions where benefits outweigh risks.

KEYWORDS

Artificial Intelligence,Risk Benefit Models, AI Challenges, AI Advantages, Generative AI

1. INTRODUCTION

While AI offers many benefits in improving efficiency, accuracy, accessibility, pattern-recognition and creating higher-paid highly skilled technology-related jobs, sustainability and quality of life, it also presents risks including invading privacy, displacing jobs, creating bias, increasing fraud/deception and weaponizing of AI.As the costs and number of risks associated with AI rise, the ability both to assess those risks and to engage all-level workers in implementing controls will become thenew competitive advantage[1].

This article addresses the risks of AI through the lens of benefit-risk analysis, based on relevant models and theories. From an analysis of the advantages and risks of AI and generative AI, this article reviews risk models and their characteristics to formulate a new paradigm for transforming AI risks into benefits. Lastly, innovative solutions based on the five dimensions of risks are elaborated to illustrate the transformational process where overall benefits outweigh the risks.

2. ADVANTAGES OF ARTIFICIAL INTELLIGENCE

As far back as the industrial revolution, significant development in technical innovation has succeeded in transforming numerous manual tasks and processes that had been in existence for decades where humans had reached the limits of physical capacity. So it is no surprise that Artificial Intelligence (AI) offers this same transformative potential for the augmentation and

David C. Wyld et al. (Eds): IoTE, CNDC, ACITY, DPPR, AIAA, NLPTA, WEST, ICSS -2023
pp. 165-179, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.132213

potential replacement of human tasks and activities within a wide range of industrial, intellectual and social applications [2]. If society and organizations managed the risks associated with the industrial revolution technologies, it is apparent that managing the challenges of AI is also within reach if leaders commit to vigilance, persistence and value-based decisions.

With new breakthroughs in algorithmic machine learning and autonomous decision-making coming at society with breakneck speed, leaders must learn how to take advantage of the new opportunities createdby AI innovation and disruption while balancing the legal, political, ethical, and governance implications of these changes in both private and public sectorsranging from finance, healthcare, manufacturing, education and retail to supply chain/logistics, utility companies and governmental agencies[2]. With its ability to analyze vast amounts of data, identify patterns and provide accurate predictions, AI can help society to solve difficult problems like cybercrimes and AI can create a better world, which includes access to education, healthcare, and clean water, and fighting climate change, poverty and hunger. By harnessing the power of AI, society can accelerate the path towards a worldwide sustainable future (Iliescu & Bustos, 2022).

In fact, from a comprehensive analysis study, AI was found to be an enabler on 134 general targets (79%) across all of the Sustainable Development Goals (SDGs) set by the United Nations, which allows the world in general to overcome certain pressing worldwide problems. In the societal and environmental areas of challenges, AI can enhance the provision of food, health, water and energy services to the population, for instance. AI can create smart and low-carbon cities encompassing a range of interconnected technologies such as electric autonomous vehicles and smart appliances that enable demand response in the electricity sectorand smart grids that match electrical demand to weather conditions [4]

By 2030, experts predict between 20 million to 50 million new AI-related jobs will be created globally (higher skill levels and ability to manage technology will be in high demand) and overall spending on technology could increase by more than 50 percent between 2015 and 2030 [5].
Studies have analyzed the societal, economic and organizational impact of this significant change, showing a changing jobs market that is predicted to focus humans on more creative and higher thinking skills in AI-support roles that pay more [6]. Generative AI is creating more types of jobs and leaders are already hiring more because of the generative AI surge[7].

Upwork Research Institute's new study surveying 1,400 U.S. business leaders—senior managers through C-suite level—confirms the positive job impact of AI.Although there may be variation in generative AI adoption of technologies such as ChatGPT and Midjourney across companies, two-thirds of top leaders agree that they will increase hiring as a result of generative AI [8]. Studies have identified the benefits of applying AI technologies to big data problems and the significant value of analytic insight and predictive capability for a number of scenarios (Rubik & Jabs, 2018). Health related studies, for example,show AI technologies can greatly support patient health-based diagnosis and predictive capability[9][10][11][12][13]. Economically, AI also promises to generate productivity gains [14]. In the legal and policy arena, the best cybersecurity solutions have AI and machine learning built into them, which creates dynamic protection that can consistently evolve to face ever-improving attacks and fraud threats that use automated scripts, such as account takeovers, credential stuffing, card fraud and fake account creation [15].

Although generative AI helps those who wish to radicalize people to create deep fake videos used online in social media, AI tools can also prevent it from succeeding. In November 2022, Intel released FakeCatcher, a cloud-based AI tool that it claims can accurately detect fake videos 96% of the time, using up to 72 different detection streams[16].

Retail and service-oriented organizations benefit from AI's ability to personalize the online experience through real-time, tailored product recommendations based on customers' purchase history, age, gender, geographic location, and other data points and continuously upgrades its capabilities by analyzing millions of customer-service interactions [17]. With the help of AI, companies can save time and resources and can create even more personalizedexperiences directly communicating with their customers, which will result in enhanced brand loyalty and lifetime relationships [17].Item recommendation is instrumental for a content provider to grow its audience and evidence shows that automated recommendations account for 35% of sales on amazon.com, 50% of initial messages sent on match.com and 80% of streamed hours on netflix.com [18].

Big Data Analytics (BDA) develops the methodological analysis of large data structurestackling data volume, velocity, variety, veracity and value issues. BDA combined with AI has the potential to transform areas of manufacturing, health and business intelligence offering advanced insights within a predictive context [19][20][21]. Organizations are increasingly deploying data visualization tools and methods to make sense of their big data structures. Because the analysis and processing of complex heterogeneous data is problematic and human perception and cognition may be limited, organizations can extract significant value, understanding and vital management information from big data via intelligent AI-based visualization tools [22][13][23].

Advantages of AI span societal (SOC), economic (ECON), ethical (ETH), political/legal/policy (POL), environmental (ENV), data (DATA), technological (TECH) and organizational/managerial (ORG) arenas along with examples are shownin Table 1.

3. RISKS OF ARTIFICIAL INTELLIGENCE

The implementation of AI technologies can present significant challenges or risks for government and organizations as the scope and depth of potential applications increase and the use of AI becomes more mainstream. Although AI-based systems are increasingly being leveraged to provide value to organizations, individuals, and society, significant attendant risks have been identified recently.

[24] divided risks into two categories including1) *Risks to fundamental rights*—flaws in the overall design of AI systems or biased data can lead to breaches in fundamental rights, including free speech, discrimination based on sex, race, religion, disability, age or sexual orientation, protection of personal data and private life and consumer protection; and 2) *Risks to safety and liability*—flaws in the design of AI technologies may present new safety risks for users related to the availability and quality of data or other problems stemming from AI and machine learning.
This research builds upon 8 AI challenge categories proposed in a public sector model, namely: 1) societal; 2) economic; 3) political, legal and policy; 4) environmental; 4) ethical; 5) data; 6) technological and implementation; and 8) organizational and managerial [13]. Public organization adoption of AI and data science presents numerous known challenges ranging from employee path dependency on embedded processes and norms, information silos, and a lack of resources, collaborative culture and technical capacities [25][26][27]. AI systems can have had some harmful side effects on communities, through effects on employment and inequality [28], privacy and safety—injury, property loss and workplace hazards [14] addictive behavior [29], fairness, bias and discrimination [30][31], human rights [12] and polarization, extremism, manipulative practices and conflict [32][33][34][18].

Table 1. Salient AI advantages by categories and examples

Categories	Examples
SOC	Worldwide productivity gains; Workers can choose what tasks they want to do; Societal increased well-being; Impact world sustainability goals of UN; Analytics and AI based predictions create better patient healthcare; Improve Customer/Organization Interface; Tackle world problems; Overcome Cultural barriers; Protect Human rights;
ECON	Increase higher-skill and higher-paying jobs; Create new job categories; Improve profitability; Optimize returns on investment;
POL	Detect deep fakes, propaganda, or spying on users; Fight cybercrimes and fraud; Stop weaponizing AI for harmful purposes; Add rules and accountability in the use of AI; Rely on European Union guidelines/rules for responsible AI
ENV	Address societal problems like climate change; Positively impact carbon footprint; Address resource scarcity;
(ETH	Build trust towards AI-based decision making; Increase explanation of decisions made by AI; improve processes relating to AI and human behavior, find balance in machine versus human value judgements; be vigilant in testing for bias/discrimination
DATA	Transform areas of manufacturing, health and business intelligence offering advanced insights within a predictive context; improve 5 V's of data; Increase transparency and reproducibility; Create reliable and sufficient data pools; Improve data integration and continuity; Create standards for data collection;
TECH	Fight adversarial attacks; Increase transparency and interpretability; Smart design of AI systems; Increase AI safety and security; gain value from big data; continue to innovate ways to decrease problems using unstructured data; integrate legacy and new systems through collaboration with vendors, governmental and educational institutions
ORG	More personalized customer experiences; Streamline repetitive and boring jobs; Use visualization to increase understanding of needs of the organizational system; Improve data sharing and collaboration; Increase AI talent through HR innovation in hiring and retraining; Create an AI-driven culture.

For example, researchers found race bias [35][36] and gender bias [37] for policing algorithms and bias against working-class and disadvantaged communities for educational assessments in the UK during COVID [38]. In addition, an AI algorithm created by Zillow was unable to accurately capture complex assessments such as estimating home values thus causing layoffs of 2,000 employees and a sell-off of its iBuying division [39]. Industry-friendly hackers also fooled Tesla's Autopilot AI program into merging into oncoming traffic and took control of the car using a video game controller [40].

To make matters worse, the high legal bar to prove either a disparate treatment or disparate impact cause of action under Title VII of the Civil Rights Act, coupled with the "black box" nature of many automated hiring systems makes the detection and redress of bias in such algorithmic systems very difficult [30].

In the intersection of privacy and human rights, AI used for target marketing/customer service may gather data that may include the user's private behaviors such as playing a certain game, smoking, watching porn or defaulting on loans)[18]. Concerns regarding online personalization range from recommendation algorithms isolating information seekers from differing viewpoints (filter bubble), radicalizing citizens' attitude towards controversial issues (polarization) or enabling malicious content [18]

Generative artificial intelligence (AI) has become widely popular, but its adoption by businesses comes with a degree of ethical risk [41]. With relatively modest amounts of data and computing power using generative AI, the average person can create a video of a world leader confessing to illegal activity leading to a constitutional crisis, a military leader saying something racially

insensitive leading to civil unrest in an area of military activity, or a corporate titan claiming that their profits are weak leading to global stock manipulation. These so called deep fakes pose a significant threat to our democracy, national security and society [32].

There are also practical issues over how accurate machine learning solutions actually are. The range of testing approaches available within machine learning is growing rapidly, and that is a good thing, but it is also driven by the evident limitations of the previous methods and the need to overcome those limitations [42].

Another important factor is the availability of suitable data. Although machine learning packages for Python and R can easily read all types of data from Excel to SQL and can perform natural language processing and process images, the speed with which machine learning solutions have been proposed has not kept pace with firms' abilities to suitably organise the internal data they have access to. Data is often held in separate silos across departments, perhaps on different systems, and perhaps with internal political and regulatory issues restricting the sharing of data. Important data might not even be recorded as data but rather kept as informal knowledge of the firm [42].

Estimates for work displacement due to automation highlight that up to a third of current work activities (between 400 million and 800 million jobs around the world) could be impacted by 2030 [5].

Another issue is the lack of skilled staff to implement these new technologies. A survey of the top 1000 firms in the United States found that their biggest concern in the implementation of AI was the readiness and ability of staff to understand and work with these new solutions [43]. Sixty-four percent of US executives and 70 percent of European leaders believe they will need to retrain, up-skill or replace a fourth of their workers due to advancing automation and digitization [44].Please see Table 2 for a summary of salient AI risk examples by categories.

4. MODELS AND THEORY

Very few comprehensive models of AI risk are available today to help managers assess and mitigate the increasing risks they face. To address this void and in light of the ever-increasing depth and breadth of AI technology challenges, the researchers summarize existing related risk models and start with an historical empirically-based risk model dealing with the adoption of information technology projects. Since the risk of IT projects failing was very high in those days, [45] developed a model to predict project risk and delineate risk factors associated with organizational IT projects, which serves as a basis for a newly proposed AI risk framework.
The McFarlan Risk Model (MRM) provided a useful and measurable approach for the diagnosis and mitigation of IT project risks with three dimensions based on 'project size', 'project structure' and 'experience with technology'. For example, if there were high costs, large numbers and levels of staff needed, increased completion time to complete and impacted many different departments, then the risk was increased. If the user department needed to change a lot of procedures and structures to meet the project requirements and users were highly resistant to changes, the project was also considered higher risk. Lastly, if the team lacked the appropriate experience with the new technologies, the firm needed to hire more experts or use outside consultants and if the experts did not work in partnership with the company, the risks expanded exponentially [45].

These three dimensions were expanded in this research to include 'resources' (not just the size of a project but how many, how well and how long resources support AI implementation), 'governance' including guidelines and structural processes put in place (not just the project

structure) and 'expertise building' or on-going capacity and commitment to train, retrain, upskill and encourage

Table 2. Salient AI risks by categories and examples

Categories	Examples
SOC	Customer/Organization Interface; World crises; Cultural barriers; Human rights; Country specific data profiles; Unrealistic expectations towards AI technology; Country specific organizational practices and insufficient knowledge on values and advantages of AI technologies.
ECON	Affordability of required computational expenses; High costs for customers; High cost and reduced profits for organizations; Wider divides in society leading to social upheaval
POL	Copyright issues; Embedded bias and discrimination by humans or technology; Injury; Governance of autonomous intelligence systems; Responsibility and accountability; reduced privacy/safety; National security threats from foreign-owned companies/governments collecting sensitive data, creating deep fakes, propaganda, or spying on users; Cybercrimes; Weaponizing AI for harmful purposes; Lack of rules and loss of accountability in the use of AI; Costly human resources still legally required to account for AI based decision; Lack of official industry standards of AI use and performance evaluation; difficult to redress bias legally.
ENV	Impact on carbon footprint, use of resources such as fossil fuels, water, electricity; impact on climate change and air quality
ETH	Lack of trust towards AI based decision making and unethical use of shared data. Responsibility and explanation of decision made by AI; processes relating to AI and human behavior, compatibility of machine versus human value judgement, moral dilemmas and AI discrimination
DATA	Lack of data to validate benefits of AI solutions; Format, quantity and quality of data; Transparency and reproducibility; Insufficient size of available data pool; Lack of data integration and continuity; Lack of standards of data collection;
TECH	Adversarial attacks; Lack of transparency and interpretability; Design of AI systems; AI safety; Specialization and expertise; Big data; Architecture issues and complexities in interpreting unstructured data; integrating legacy and new systems
ORG	Realism of AI; Better understanding of needs of the organizational system; Organizational resistance to data sharing and collaboration; Lack of inhouse and interdisciplinary AI talent; Threat of workforce layoffs and retraining; Lack of strategy for AI development; Embedded processes and norms; lack of resources; information silos

employee development (not just relying on the current experience of the workers or hiring outsiders).

In their Risk Assessment Framework on implementing enterprise resource planning projects, risks from 'external engagement', program management, work stream and work package levels across technical, operational, business and organizational categories were successfully mitigated using on-going risk controls [46].

The specific content of corporate governance guidelines and policies is an important variable to any risk mitigating model. Using a comparison between Belgium (weaker country guidelines) and Australia (stronger governance guidelines), researchers found significantly more developed risk management and internal control systems in Australian companies versus Belgian companies overall [47].

In addition to strong governance guidelines, impactful 'governance' also dictates that an entire risk management structure be in place. For example, the structure of risk control may include 4

phases of identifying, assessing, mitigating and monitoring risks [48] or 4 specific functions — govern, map, measure and manage — to help organizations address the risks of AI systems [49]. In this research, the 'governance' dimension of risk focuses on a broader range of AI policies, controls and structure. Using a comprehensive AI risk management system can maximize the benefits of AI technologies while reducing the likelihood of negative impacts to individuals, groups, communities, organizations and society" [49].

Companies that succeed in turning risks into results will create competitive advantage through more efficient deployment of scarce resources, better decision-making and reduced exposure to negative events. Leaders who apply a broad "risk lens" to their businesses will be most effective in transforming AI[50].

In a survey of 576 companies and a review of 2, 750 company/analyst reports, the researchers found that financial performance is highly correlated with the level of integration and coordination across risk, control and compliance functions. Effectively harnessing technology to support risk management is the greatest weakness or opportunity for most organizations [50].
To improve the predictive ability of McFarlan's model, the Extended McFarlan Risk Model (EMRM) was developed to differentiate between project success and failure, adding an organizational' culture' dimension [51]. The 'culture' attributes were quantitatively measured including users' practices, users' attitudes, company working practices, organizational polices such as information technology policy and data flow practices, internal and external communication practices in the organization, openness to change and cross-functional coordination. If the corporate culture was lacking, the implementation risk increased [51].In this research the dimension of 'culture' was widened to encompass an 'AI-driven culture' with communication, work practices, policies, change management, innovation and cross-disciplines/perspectives which fully integrate AI throughout the firm for triple bottom line goals of profits, people and the planet[52].

The new risk model proposed is also based on the use of social cognitive theory which views people as active agents who can both influence and be influenced by their environment [53].That is, people learn or are transformed by observing and interacting with others. For example, in her research, success in using cloud computer services was dependent on factors such as 'external engagement' with people who were using the technology and workers' personal attitude towards risk and innovation [54].

The researchers propose a new dimension of 'transformational collaboration'(not just 'external engagement' outside the firm) across boundaries such as disciplines, departments, divisions, firms, industries, societies and countries which emphasizes that a person's attitude and collaboration with others who are using new technologies such as AI or those who have different backgrounds can help increase technology implementation success.

Combining this systematic literature analysis with related risk models and theory, this research adds value to the field of AI by synthesizing and creating an integratedmodel for managing risk designed particularly for AI adoption and implementation. Using [45] as a theoretical lens, the updated McFarlan model [46] for the context of culture and social cognition theory's [53] contributions of personal attitude and engagementfrom previous risk research [46],theAI Transformation Risk Model proposes an innovative model that expands to 5 broader dimensions of 'resources' (RES),'governance' (GOV), 'expertise building' (EXP), 'AI driven culture' (AI CUL) and 'transformational collaboration' (TR COL) that has the potential tochangeAI risks into benefits as presented in Figure1.

5. SOLUTIONS

Companies play a critical role in responsibly adopting AI and generative AI, and integrating these tools in ways that enhance, not diminish, the working experience of their employees and their customers. This comes back to ensuring the responsible use of AI in maintaining accuracy, safety, honesty, empowerment, and sustainability, while mitigating risks and eliminating biased outcomes. Corporate commitment should extend beyond immediate corporate interests, encompassing broader societal responsibilities and ethical AI practices as well [41].
Leaders remain in need of conceptual, technical and institutional mechanisms to assess how to achieve accountability for the harmful consequences of data-driven algorithmic systems—mechanisms that address both whom to hold accountable and how to hold them accountable [56].

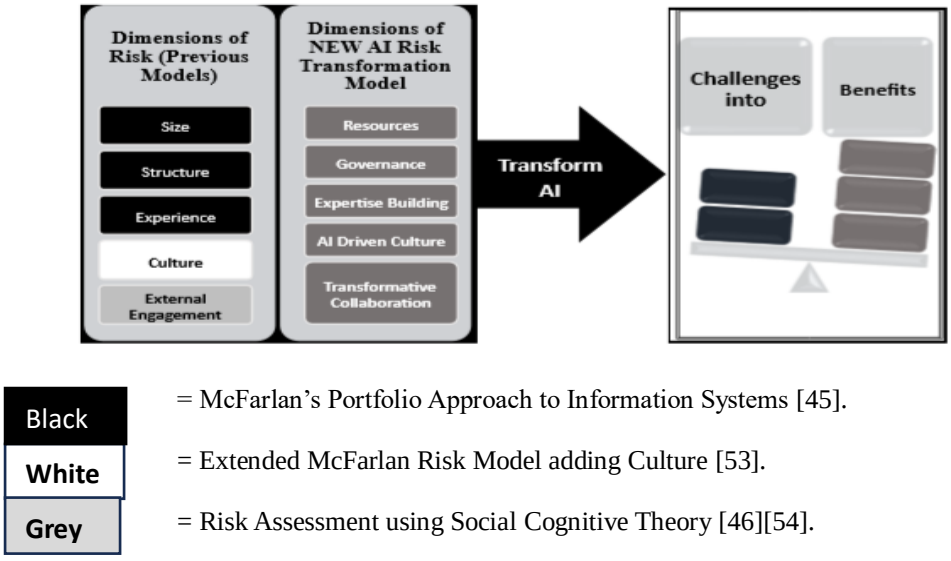

Figure 1. AI Risk Transformation Model: Transform AI challenges into benefits

In this section, the authors give examples of innovative solutions that can transform risks into benefits in each of the model's 5 risk dimensions. Under the dimension of 'resources', innovative solutions such as developing policies to assure AI will be directed at 'humanness' and common good and building inclusive, decentralized intelligent digital networks 'imbued with empathy' will help leaders meet social and ethical responsibilities while using new AI technologies[57].
Another resource-related solution is to alter organizational, economic and political systems to better help humans 'race with the robots' and reorganize these systems toward the goal of expanding humans' capacities and capabilities which will lead to stronger human/AI collaboration which in turn lowers the risk of compromising humans in the face of AI [57].
Beyond retraining, a range of resource policies can help, including unemployment insurance, public assistance in finding work and portable benefits that follow workers between jobs [5].
Increasing human/AI collaboration and human interventions in decision-making particularly in high risk situations like finance, healthcare and hiring practices will ensure trust, maintain transparency and decrease harms to people [41].

Under the 'governance' dimension, comprehensive AI lifecycle governance where policies and procedures are described and enforced during the design, development, deployment, and monitoring phases of an AI system lifecycle can increase transparency. AI factsheets are one example that captures model metadata across the AI lifecycle automatically, making enterprise validation and external regulation easier to monitor [58].

To increase fair and unbiased 'governance', create "an auditing imperative" for algorithmic hiring systems that mandates regular internal and external audits of automated systems, as well as detailed record-keeping for all job applications. There is precedent in other areas of law such as the Occupational Safety and Health Administration (OSHA) audits in labor law or the Sarbanes-Oxley Act audits in securities law, these steps ensure trust and confidence in AI- automated systems[30].

Other effective 'governance' solutions are to conduct your own bias, explainability and robustness assessments to protect privacy and harmful outputs and security assessments to help identify and decrease threats such as cybercrimes for the organization. Also prioritize the responsible use of AI and generative AI by ensuring it is accurate, safe, honest, empowering and sustainable by using "zero or first party data, keeping data fresh and well labeled, ensuring there's a human in the loop, testing and re-testing and keeping feedback close" [41].

Base 'governance' on protecting society from harm and unwanted negative consequences for the firm by following stricter regulations, guidelines & certification standards such as those recommended by the European Union[24].

To increase the second dimension of 'expertise building', train or retrain workers in new fields such as digital forensics, cybercrime specialists, data analysts and AI governance specialists, use Protective Optimization Technologies (POTs) to both explore the effects that algorithms and optimization systems have on our society, and the design of countermeasures to contest their negative effects [59].

Widen educational opportunities to step into new AI jobs and learn new skills throughout the society (throughout the job and educational systems available or create new ones), build talent in technical competencies like AI engineering and enterprise architecture and empower people across the organization to work effectively with AI-infused processes[57].

These solutions collectively can: 1) create new jobs, tools and skills that help detect deep fakes, harmful patterns in the digital arena and make workers more productive; 2) bring structure and strategies to AI tool and skill use that actually help not harm society and 3) pairing more educational/training options with funding (grants, scholarships, internships) will increase the availability of skills throughout the organization and society and 4) increase firm efficiency and give workers opportunities to gain higher-skilled higher-paid jobs.

In the arena of creating an 'AI driven culture', use AI-based visualization tools [13][23],leverage open-source and user-provided data, be transparent when AI has created content using watermarks or in-app messages, cite the sources from where the model is creating content, explaining why the AI gave the response it did, highlighting uncertainty and creating guardrails, extend the commitment to using AI tools and technologies beyond immediate corporate interests, helping meet broader societal responsibilities and ethical AI practices and prevent some sensitive and risk heavy tasks from being fully automated [41].

These innovative solutions for developing an 'AI driven culture' willhave important benefits such as extracting significant insights from big data and allowing workers to understand the impact of AI directly, increasing transparency, accuracy and honesty about data uses and sources in the organization and broadly across societies, mitigating risks and eliminating biased outcomes while protecting sensitive data processes and uses [41].

In the final dimension of risk, 'transformational collaboration' solutions could entail: 1) working collaboratively to minimize the size of AI models[41], 2) collaborating with vendors and partners

to create better data and train on models with large amounts of high-quality customer relationship management or other generated data; or 3) increasing collaboration and risk prevention and control discussions across departments, disciplines, companies, countries and societies to solve big complicated or wicked problems [57].

Working collaboratively can produce outstanding outcomes while lowering threats such as reducing the carbon footprint because less computation is required in smaller models, which means less energy (water and electricity) consumption from data centers and carbon emissions, higher quality data helps to maximize data accuracy, reliability and value, and working across boundaries and perspectives can facilitate the innovation of widely accepted approaches aimed at tackling wicked problems such as climate change or world health and improves risk control over complex human-digital networks worldwide[57]. Please see Table 3 for a sampling of innovative solutions in the five dimensions of risk that are poised to transform AI risks into benefits.

Table 3. Innovative sample solutions by risk dimensions transforming AI risks into benefits

Dimensions	Innovative Sample Solutions	Transforming AI Risks into Benefits
RES	• Prioritize people by reorganizing organizational, economic and political systems towards the goal of expanding humans' capacities and capabilities • Increase human/AI collaboration and human interventions particularly in high risk situations.	• Helps humans 'race with the robots' and staunches AI trends that would compromise human relevance • Improves trust and removes harms in finance, healthcare or hiring.
GOV	• Build comprehensive AI lifecycle governance where policies and procedures are described and enforced during the design, development, deployment, and monitoring of an AI system. • Create AI factsheets to capture AI metadata across the model lifecycle automatically, • Create "an auditing imperative" mandate for both internal and external audits of automated hiring systems • Keep and audit records of job applications. • Use AI tools for security and threat assessments • Conduct bias, explainability and robustness assessments • Base governance on protecting society from harm and unwanted negative consequences for the firm by following stricter regulations, guidelines & certification standards recommended by the European Union • Test and re-test and get feedback close to source	• Builds trust in AI through transparency • Facilitates enterprise validation or external regulation • Builds trust and confidence in decision making and has been used in other areas of law, such as Occupational Safety and Health Administration (OSHA) audits in labor law or the Sarbanes-Oxley Act audit requirements in securities law • Helps organizations identify vulnerabilities that may be exploited by bad actors such as cybercrimes • Protects privacy and harmful outputs • Self-regulation decreases risk of harms • Ensure data is high quality, accurate and has value
EXP	• Train or retrain workers in new fields such as digital forensics, cybercrime specialists, data analysts and AI governance specialists • Protective Optimization Technologies (POTs) systematizes the use of technologies as tools to both explore the effects that algorithms	• New jobs, tools and skills can help detect deep fakes, harmful patterns in the digital arena and make workers more productive • Brings structure and strategies to AI tool and skill use that actually help not harm society

	and optimization systems have on our society, and the design of countermeasures to contest their negative effects • Widen educational opportunities to step into new AI jobs and learn new skills • Building talent in technical competencies like AI engineering and enterprise architecture and training people across the organization to work effectively with AI-infused processes.	• More educational options along with funding (grants, scholarships, internships) will increase the availability of skills throughout the organization and society • Benefits the organization by increasing efficiency and gives workers higher skilled higher paid jobs
AI CUL	• Use AI-based visualization tools • Leverage open-source and user-provided data. • When autonomously delivering outputs, be transparent that an AI has created the content using watermarks on the content or through in-app messaging • Responsibly use AI in maintaining accuracy, safety, honesty, empowerment and sustainability by citing the sources from where the model is creating content, explaining why the AI gave the response it did, highlighting uncertainty and creating guardrails • Extend commitment beyond immediate corporate interests, encompassing broader societal responsibilities and ethical AI practices • Prevent some tasks from being fully automated	• Extracts significant value and key management information from big data • Ensures honesty about the use and sources of data • Increases transparency • Mitigates risks and eliminates biased outcomes. • Improves the accuracy of data by being transparent and in the organization and broadly across societies. • Human intervention may protect certain risky & sensitive data processes and uses.
TRAN COL	• Work collaboratively to minimize the size of AI models • Collaborate with vendors and partners to create better data and train on models with large amounts of high-quality customer relationship management or other generated data. • Collaborate across departments, disciplines, companies, countries and societies to solve big complicated or wicked problems.	• Reduces carbon footprint because less computation is required, which means less energy consumption from data centers and carbon emissions • Maximizes data accuracy, reliability and value • Facilitates the innovation of widely accepted approaches aimed at tackling wicked problems and maintaining control over complex human-digital networks worldwide

6. CONCLUSIONS

This research has systematically reviewed the most salient advantages and risks of AI and generative AI in the current literature and categorized them by their impact on society.From this analysis, examples of these challenges and benefits are categorized into societal, economic, ethical political/legal/policy, environmental, data, technological and organizational/managerial arenas based on the research.

Next, relevant risk models and theories, based upon risk models in IT project management, cloud computing implementation, enterprise resource management and social cognitive theory applied

from psychology were compared and synthesized into the context of AI resulting in the newly proposed "AI Risk Transformation Model." The model presents a comprehensive framework tailored to handle the ever-increasing risks of the AI environment.

The 'AI Risk Transformation Model' adds to the research by expanding the previous categories of risks (such as 'project size' transforming into 'resources','culture' changing to 'AI-driven culture' and 'external engagement' widening to 'transformational collaboration' both internally and externally to the organization) in light of AI-related challenges found in the systematic review of the literature. Lastly, the research adds to the body of knowledge by presenting samples of pragmatic and innovative solutions in each of the AI and generative AI risk dimensions that highlight how risk transformation can occur in the context of the latest AI technologies.

REFERENCES

[1] Cheatham, B., Javanmardian, K. & Hamid Samandari, H. (2019, April 26) "Confronting the risks of artificial intelligence." McKinsey Quarterly, Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence#/

[2] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Janssen, M. et al. (2021). "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy." International Journal of Information Management. Vol. 57, doi:10.1016/j.ijinfomgt.2019.08.002

[3] Iliescu, A.B. & Bustos, A. (2023, March 29) "How using AI can benefit society?" AI for Good, Available:https://aiforgood.itu.int/howusingaicanbenefitsociety/#:~:text=With%20its%20ability%20to%20analyze,change%2C%20poverty%2C%20and%20hunger.

[4] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M. & Nerini, F. F. (2020) "The role of artificial intelligence in achieving the Sustainable Development Goals." National Communications, Vol. 11, pp1-233.

[5] Manyika, J., Lund, S. Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R. & Sanghvi, S. (2017, November 28) "Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages?' McKinsey Global Institute, Available: https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages

[6] Jonsson, A. & Svensson, V. (2016) "Systematic lead time analysis." Chalmers University of Technology,Available:http://www.publications.lib.chalmers.se/records/fulltext/238746/238746.pdf

[7] Curry, Rachel (2023, August 10) "How AI can help create jobs for humans, not just automate them." Technology Executive Council, Available: https://www.cnbc.com/2023/08/10/how-ai-can-help-create-jobs-for-humans-not-just-automate-them.html

[8] Monahan, Kelly (2023, June 27) "Generative AI disconnect." Upwork Available: https://www.upwork.com/blog/generative-ai-disconnect

[9] Beregi, J.P., Zins, M., Masson, J.P., Cart, P., Bartoli, J.M., Silberman, B., Boudghene, F. & Meder, J.F. (2018) "Radiology and artificial intelligence: An opportunity for our specialty," Diagnostic and Interventional Imaging, Vol. 99, No. 11, pp677-678, doi:10.1016/j.diii.2018.11.002

[10] Bergquist, T., Yan, Y., Schaffter, T., Yu, T., Pejaver, V., Hammarlund, N., Prosser, J., Guinney, J. & Mooney, S. (2020) "Piloting a model-to-data approach to enable predictive analytics in health care through patient mortality prediction." Journal of the American Medical Informatics Association : JAMIA, Vol. 27, No. 9, pp1393-1400.

[11] Harada, T., Miyagami, T., Kunitomo, K. & Shimizu, T. (2021, August 10) "Clinical Decision Support Systems for diagnosis in primary care: A scoping review." International Journal of Environmental Research and Public Health, Vol. 18, doi:10.3390/ijerph18168435

[12] Metzger, M. & Donahoe, E. (2019) "Artificial Intelligence and human rights." Journal of Democracy, Vol. 30, No. 2, pp115-126.

[13] Zheng, Y. & Hu, X. (2020) "Healthcare predictive analytics for disease progression: a longitudinal data fusion approach." Journal of Intelligent Information Systems, Vol. 55, pp351–369.

[14] OECD correspondent (2019) Artificial intelligence in society. OECD Publishing. doi:10.1787/eedfee77-en.

[15] Data Dome Correspondent (2021, September 18), "How AI is used in fraud detection," Data Dome, Available: https://datadome.co/learning-center/ai-fraud-detection/

[16] Gregory, J. G. (2023, Sept 23) "How Data Scientists Fight Deep Fakes in Cyberspace." Nutanix,https://www.nutanix.com/theforecastbynutanix/technology/
[17] Zulaikha, S., Mohamed, H., Kurniawati, M. Rusgianto, S. & Rusmita. S.A. (2021) "Customer predictive analytics using Artificial Intelligence," The Singapore Economic Review, Vol. 66, No. 2, pp1-12.
[18] Stoica, A.A. & Chaintreau, A. (2019) "Hegemony in social media and the effect of recommendations."The Web Conference 2019, Vol. 2, pp575-580.
[19] Abarca-Alvarez, F.J., Campos-Sanchez, F.S., Reinoso-Bellido, R. (2018) "Demographic and dwelling models by artificial intelligence: urban renewal opportunities in Spanish Coast." International Journal of Sustainable Development and Planning, Vol. 13, No. 7, pp941-953.
[20] Shukla, N., Tiwari, M. K. & Beydoun, G. (2019) "Next generation smart manufacturing and service systems using big data analytics," Computers & Industrial Engineering, Vol. 128, pp905-910
[21] Spanaki, K., Gürgüç, Z., Adams, R. & Mulligan, C. (2018), "Data supply chain (DSC): research synthesis and future directions," International Journal of Production Research, Vol. 56, No.13, pp4447-4466.
[22] Olshannikova, E., Ometov, A., Koucheryavy, Y. & Olsson, T. (2015) "Visualizing Big Data with augmented and virtual reality: Challenges and research agenda," Journal of Big Data, Vol. 2, pp22-49.
[23] Zhong, R. Y., Xu, C., Chen, C. & Huang, G.Q. (2015) "Big Data Analytics for physical internet-based intelligent manufacturing shop floors," International Journal of Production Research, pp1-14.
[24] Ciucci, M. & Gouardères, F. (2020, April) "Briefing, requested by the ITRE Committee, European Parliament, white paper on Artificial Intelligence", PE 648.773, Policy Department for Economic, Scientific and Quality of Life Policies, Directorate-General for Internal Policies, doi: 10.2861/614816
[25] Mergel, I. (2018) "Open innovation in the public sector: Drivers and barriers for the adoption of challenges," Public Managerial Review, Vol. 20, pp726–745.
[26] Mikhaylov, S. J., Esteve, M. & Campion, A. (2018). "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration." Philosophical transactions. Series A, Mathematical, physical, and engineering sciences Vol. 376, No. 2128, e20170357. doi:10.1098/rsta.2017.0357.
[27] Eggers, W. & Bellman, J. (2015) "The journey to government's digital transformation," Deloitte. Available: https://www2.deloitte.com/uk/en/pages/public-sector/articles/the-journey-to-governments-digital-transformation.html
[28] Korinek, A. & Stiglitz, J. E. (2020, January 23) "Artificial Intelligence and its implications for income distribution and unemployment." Chapter 14, pp349-390 In The Economics of Artificial Intelligence: An Agenda. Ajay Agrawal, Joshua Gans and Avi Goldfarb (Editors). University of Chicago Press, Chicago Scholarship Online, 2020. doi:10.7208/chicago/9780226613475.003.0014
[29] Andreassen, CS. (2015) Online social network site addiction: A comprehensive review. Current Addiction Reports. Vol. 2, No. 2, pp175–184.
[30] Ajunwa, I. (2021) An auditing imperative for automated hiring, Harvard Journal of Law & Technology Vol. 34, No. 2, pp1-66.
[31] Barocas, S., Hardt, M. & Narayanan, A. (2018) Fairness and Machine Learning: Limitations and Opportunities. Available: http://fairmlbook.org
[32] Agarwal, S., Farid, H., Gu, Y. , He, M., Nagano, K. & Li, H. (2019) "Protecting World Leaders Against Deep Fakes." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. IEEE, Long Beach, CA, pp38-45.
[33] Kreps, S., McCain, R. M. & Brundage, M. (2020, September 24). "All the news that's fit to fabricate: AI-Generated text as a tool of media misinformation. Journal of Experimental Political Science, pp1–14.
[34] Ledwich, M. & Zaitsev, A. (2020) "Algorithmic extremism: Examining YouTube's rabbit hole of radicalization." First Monday, Vol. 25, No. 3, doi:10.5210/fm.v25i3.10419.
[35] Larson, J. Mattu, S., Kirchner, L. & Angwiny, J. (2016, May 23) "How we analyzed the COMPAS recidivism algorithm," ProPublica. Available: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
[36] Selbst, A.D. (2017, February 25) "Disparate impact in big data policing," Georgia Law Review, Vol. 52, No. 1, pp109–195.

[37] Dastin, J. (2018, October 10) "Amazon scraps secret AI recruiting tool that showed bias against women," Rueters, Oct. 2018. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[38] Hao, K. (2020, August 20) "The UK exam debacle reminds us that algorithms can't fix broken systems," MIT Technology Review, Available: https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/

[39] Cook, J. (2021, Nov. 3) "Why the ibuying algorithms failed Zillow, and what it says about the business world's love affair with AI," GeekWire, Available: https://www.geekwire.com/2021/ibuying-algorithms-failed-zillow-says-business-worlds-love-affair-ai/

[40] Rapier, G. (2019, April 1) "Hackers steered a Tesla into oncoming traffic by placing 3 small stickers on the road," Business Insider, Available: https://www.businessinsider.com/teslahackers-steer-into-oncoming-traffic-with-stickers-on-the-road-2019-4

[41] Baxter, K. & Schlesinger, Y. (2023, June 6) "Managing the risks of generative AI." Harvard Business Review, Available: https://hbr.org/2023/06/managing-the-risks-of-generative-ai

[42] Aziz, S. & Dowling, M. (2019) "Machine learning and AI for Risk Management," in Chapter 3 pp33-50, Disrupting Finance: FinTech and Strategy in the 21st Century. Theo Lynn, John G. Mooney, Perangelo Rosati and Mark Cummins, (Eds.) Palgrave Studies in Digital Business and Enabling Technologies. Palgrave Pivot. doi:10.1007/978-3-030-02330-0_3

[43] Wilson, H. J., Daugherty, P. & Bianzino, N. (2017) "The jobs that artificial intelligence will create." MIT Sloan Management Review, Vol. 58, No. 4, pp14–16.

[44] Illanes, P., Lund, S., Mourshed, M., Rutherford, S. & Tyreman, M. (2018, January 22) Retraining and reskilling workers in the age of automation, McKinsey Global Institute, Available: https://www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation?cid=other-eml-ofl-mip mck&hlkid=23d2fe2e5f04408ca17297de4d9e61cc&hctky=14997643&hdpid=de160046-2a99-4422-833a-a657c3a21cf0

[45] McFarlan, F. W. (1981) "Portfolio approach to information systems," Harvard Business Review, Sept – Oct, pp142-150.

[46] Dey, P. K., Clegg, B. & Cheffi. W. (2011) "Risk management in enterprise resource planning implementation: a new risk assessment framework," Production Planning & Control, doi:10.1080/09537287.2011.597038

[47] Sarens, G. & Christopher, J. (2010). "The association between corporate governance guidelines and risk management and internal control practices: Evidence from a comparative study." Managerial Auditing Journal, Vol. 25, No. 4, pp288-308.

[48] Ganesh, A. D. & Kalpana, P. (2022) "Future of artificial intelligence and its influence on supply chain risk management – A systematic review," Computers & Industrial Engineering, Vol. 169, e108206.

[49] Locascio, L.A. (2023, January 26) NIST Risk Management Framework Aims to Improve Trustworthiness of Artificial Intelligence: New guidance seeks to cultivate trust in AI technologies and promote AI innovation while mitigating risk. pp1-48, National Institute of Standards and Technology (NIST), U.S. Department of Commerce, doi:10.6028/NIST.AI.100-1

[50] van Kessel, P., Layman, J., Blackmore, J., Burnet, I. & Haroda, S. (2013)."Turning risk into results: How leading companies use risk management to fuel better performance." Ernst and Young Global Limited, pp1-16.

[51] Yip, A, Brehm, L. & Wagner, C. (2017) "Revisiting McFarlan's Risk Model for IS implementation success - Does culture matter?" in Association for Information Systems 2017 Proceedings of the 28th Australasian Conference on Information Systems, Hobart, Australia, Vol. 32, May 12. Available: https://aisel.aisnet.org/acis2017/32/

[52] Elkington, J. (1994). "Towards the sustainable corporation: Win-Win-Win business strategies for sustainable development," California Management Review, Vol. 36, No. 2, pp90–100.

[53] Bandura, A. (1989). Human agency in social cognitive theory. American psychologist, Vol. 44, No. 9, pp1175-1186.

[54] Ratten, V. (2016) "Continuance use intention of cloud computing: Innovativeness and creativity perspectives," Journal of Business Research, Vol. 69, No. 5, pp1737-1740.

[55] Ali, A., Warren, D. & Mathiassen, L. (2017) "Cloud-based business services innovation: A risk management model" International Journal of Information Management, Vol. 37, No. 6, pp639-649, doi: 10.1016/j.ijinfomgt.2017.05.008

[56] Cooper, A.F., Moss, E., Laufer, B. & Nissenbaum, H. (2022) "Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning," in 2022 ACM Conference on Fairness, Accountability, and Transparency Proceedings, June 21-24, 2022, Seoul, Republic of Korea, ACM, New York, NY, USA, pp864–876. doi:10.1145/ 3531146.3533150

[57] Anderson, J., Rainie, L. & Luchsinger, A. (2018, December 10) "Artificial Intelligence and the future of humans" Pew Research Center, pp1-116.

[58] Sabhlok, S. (2022, January) "Introducing AI factsheets on cloud pak for data as a service: Automate collection of model facts across the AI Lifecycle," Community IBM, Available: https://community.ibm.com/community/user/datascience/blogs/shashanksabhlok/2022/01/23/ai-factsheets-on-cloud-pak-for-data-as-a-service-a

[59] Kulynych, B., Overdorf, R., Troncoso, C. & Gürses, S. (2020) POTs: Protective Optimization Technologies. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. January 27-30, 2020, Barcelona, Spain, ACM, New York, NY, doi: 10.1145/3351095.3372853

AUTHORS

Professor Richard A. Fulton (M.S., Illinois State University)has taught full time computer science and information systems courses at Troy University – e campus for the past 15 years and previously at Illinois State University. His articles have been published in The Journal of Technology Research,The Journal of Scientific Information on Political Theory, Developments in Business Simulations and Experiential Learning, and the International Journal of Innovation, Technology and Management.

Dr. Diane J. Fulton (Ph.D., University of Tennesee-Knoxville) is Emeritus Professor of Management at Clayton State of University, located in Morrow, Georgia. Her research interests include advanced technologies, innovations and online teaching tools. She has Published several books, book chapters, and numerous articles in academic journals, including California Management Review, Planning Review, Journal of Small Business Management, International Journal of Management Education and Entrepreneurship Theory and Practice.

Susan Kaplan (BAS,MAS, University of Minnesota-Duluth) is Executive Vice President and Chief Management Officer of Modal Technology Corporation, a high-tech firm located in Minneapolis, Minnesota that offers new and proven solutions for artificial intelligence and machine learning are based on the modal interval arithmetic Ms.Kaplan was Founder are President for Quality Management Solutions. She led and restructured organizations in healthcare government, manufacturing, and service sectors to improve profitability. She is the author of the Grant Writing Process.

Nathan Hayes, CEO and Founder Modal Technology Corporation is an entrepreneur, mathematician, and software architect with more than 20 years of experience working in these combined fields,, Hayes specializes in the applied science of modal interval analysis to the fields of artificial intelligence, machine learning, and high performance computing.