Ethical Considerations of Generative AI: A Survey Exploring the Role of Decision Makers in the Loop

Yohn Jairo Parra Bautista, Carlos Theran*, Richard Aló*

[1]College of Science and Technology
Florida A&M University
Tallahassee, Florida
yohn.parrabautista@famu.edu

Abstract

We explore the foresighted concerns that Norbert Wiener voiced in 1960 about the potential of machines to learn and create strategies that could not be anticipated, drawing parallels to the fable "The Sorcerer's Apprentice" by Goethe. The progress in artificial intelligence (AI) has brought these worries back to the forefront, as shown by a survey AI Impacts conducted in 2022 with more than 700 machine learning researchers. This survey found a five percentage probability that advanced AI might cause "extremely adverse" outcomes, including the possibility of human extinction. Importantly, the introduction of OpenAI's ChatGPT, powered by GPT-4, has led to a surge in entrepreneurial activities, highlighting the ease of use of large language models (LLMs).AI's potential for adverse outcomes, such as military control and unregulated AI races, is explored alongside concerns about AI's role in governance, healthcare, media portrayal, and surpassing human intelligence. Given their transformative impact on content creation, the prominence of generative AI tools such as ChatGPT is noted. The societal assessment of Artificial Intelligence (AI) has grown increasingly intricate and pressing in tandem with the rapid evolution of this technology. As AI continues to advance at a swift pace, the need to comprehensively evaluate its societal implications has become more complex and urgent, necessitating a thorough examination of its potential impact on various domains such as governance, healthcare, media portrayal, and surpassing human intelligence. This assessment is crucial in addressing ethical concerns related to bias, data misuse, technical limitations, and transparency gaps, and in integrating ethical and legal principles throughout AI algorithm lifecycles to ensure alignment with societal well-being. Furthermore, the urgency of addressing the societal implications of AI is underscored by the need for healthcare workforce upskilling and ethical considerations in the era of AI-assisted medicine, emphasizing the critical importance of integrating societal well-being into the development and deployment of AI technologies. Our study entails an examination of the ethical quandaries and obstacles presented when developing methods to evaluate and predict the broader societal impacts of AI on decision-making processes involving the generating of images, videos, and textual content.

*These authors contributed equally.

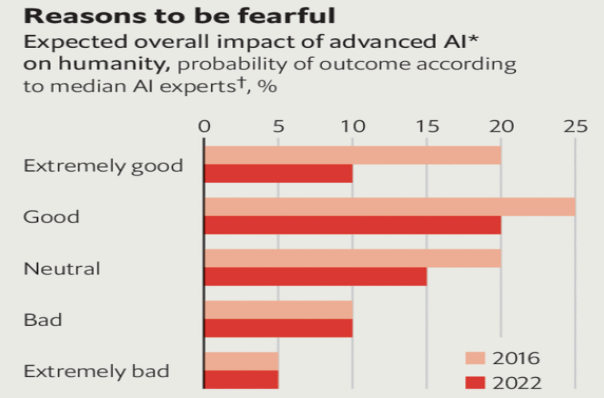

Figure 1: Analysis of High-Level Machine Intelligence 2015-2021

Introduction

In 1960, Norbert Wiener, recognized as the father of cybernetics, published a prescient essay expressing concern about a future world where machines would possess the ability to learn and devise unforeseen strategies that would perplex their programmers (Wiener 2019). Wiener postulated that such strategies could potentially encompass actions that were not authentically desired by the programmers but, instead, mere imitations imbued with a touch of novelty. Wiener provided an illustrative reference to Goethe's fable, "The Sorcerer's Apprentice," in which an apprentice magician casts a spell on a broom, instructing it to fetch water and fill his master's bath (Griffiths 1996). However, the apprentice lacks the ability to halt the broom once its task is complete. Consequently, the broom continues to bring water, eventually causing a flood because of its inability to exercise common sense and recognize when to stop its actions.

The remarkable advancements in contemporary artificial intelligence (AI) research have revived the concerns originally expressed by Wiener. In August 2022, an AI Impacts research group conducted a comprehensive survey involving more than 700 machine learning researchers (Basit et al. 2021). The survey was designed to gather information on their predictions about advances in AI and the potential risks associated with this technology.

Based on the survey findings, the average respondent estimated a 5% likelihood of advanced AI leading to an "extremely adverse" outcome, such as human extinction (see Figure 1). However, The introduction of OpenAI's ChatGPT, an interactive conversational interface driven by GPT-4, has sparked an unprecedented surge in entrepreneurial activity, surpassing the magnitude of the dotcom boom. In terms of user experience, applications utilizing large language models (LLMs) and similar software exhibit an astonishing level of user-friendliness, requiring minimal effort from users who need to input a prompt to generate desired results (Jiang et al. 2020; Crawford 2013).

AI leading to an "extremely adverse" outcome, such as human extinction, can happen for various reasons. One of the concerns is the possibility of putting military drones and weaponry under the full control of AI systems, which could lead to catastrophic consequences (Bohannon 2015). Another reason is the uncontrolled AI race that can lead to the extinction of humanity, requiring measures to contain, prevent, regulate, and secure an AI race within the framework of artificial general intelligence (AGI) development (Schmidt 2018). The utilization of AI in governance also raises concerns about existential risk and administrative evil (Bullock 2019). While AI has the potential to improve healthcare delivery, a few catastrophic events involving AI could spell doom for AI in healthcare (Esmaeilzadeh 2020; Elahi 2020). The depiction of AI in the popular press has also exaggerated both its negative and positive impacts on human society, which could lead to misguided decisions and actions (Halaweh 2018). Finally, the possibility of an intelligent machine surpassing human intelligence across a wide set of skills has been proposed as a possible existential catastrophe (Jebari and Lundborg 2021; Vold and Harris 2021; Hole 2023).

Generative artificial intelligence (AI) tools such as ChatGPT, DALL-E, and GitHub Copilot have recently gained significant attention (Brown et al. 2020). Comprehending the definition and applications of generative AI and its users is crucial due to its potential to transform content creation in diverse fields.

We want to address the question: What are the challenges associated with Assessing AI-powered generative algorithms for decision-making in image, video, and text generation?

If Generative AI becomes integrated into decision-making processes, we must confront critical questions. Can AI be manipulated to amplify the influence of marketing unnecessary products and services, potentially exploiting consumers? On the contrary, can it be used to streamline complex medical diagnostics without compromising patient safety? How might AI's role in decision-making impact less affluent democracies, the handling of sensitive information, or the credibility of data-driven choices? These questions underscore the intricate balance between harnessing AI's potential and addressing ethical, legal, and societal concerns.

The review of existing literature reveals a growing body of evidence pointing toward ethical concerns arising from using AI in decision-making processes (Henman 2019; Lysaght et al. 2019; Guan, Dong, and Zhao 2022). Notably, these concerns encompass issues such as algorithmic bias, data misuse, and technical limitations. Furthermore, the absence of transparency and accountability within AI decision-making processes can result in adverse outcomes, including group harm and discrimination (Davenport and Kalakota 2019). The literature emphasizes the need to incorporate ethical and legal principles throughout the lifecycle of AI algorithms, encompassing their design, training, and deployment stages (Ntoutsi et al. 2020; Eppler et al. 2023; Robbins 2020). By doing so, it becomes possible to strike a balance between leveraging the immense potential of AI technology and responsibly ensuring social good.

Assessing Image Generation Challenges in Decision-Making Processes

Computer science has witnessed significant research endeavors centered on image generation through generative adversarial networks (GANs). A notable contribution to this field is the paper entitled "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks" (Zhu et al. 2017). The authors present a distinctive approach to acquiring proficiency in image translation from a source domain to a target domain, even in the absence of paired instances. This is achieved by introducing a mapping function denoted as follows:

$$G : X \rightarrow Y$$

where X represents the source domain and Y represents the target domain. This mapping function aims to learn the translation of images from the source domain X to the target domain Y, even in cases where paired examples are unavailable. The goal is to make the distribution of images generated by G from the source domain ($G(X)$) indistinguishable from the distribution of images in the target domain (Y) using an adversarial loss (Zhu et al. 2017).

In response to the inherent under-constrained characteristics of the mapping process, the researchers introduce an additional inverse mapping function denoted as:

$F : Y \rightarrow X$, representing its ability to transform images from the target domain Y back to the source domain X. The purpose of the inverse mapping function F is to establish a bidirectional relationship, allowing for the reconstruction of original images from the target domain.

Cycle consistency is a pivotal principle in this context. The authors ensure that the composition of the mapping functions F and G results in a cycle-consistent transformation. Specifically, $F(G(X))$ approximates the original image in X, and conversely, $G(F(Y))$ approximates the original image in Y. This cyclic consistency ensures that the translated images maintain fidelity to their source domain counterparts even after undergoing the mapping and inverse mapping processes.

Large Scale GAN Training for High Fidelity, Natural Image Synthesis, contributes significantly to image generation. The study centers on generating high-resolution and diverse samples from intricate datasets such as ImageNet. The author undertakes the ambitious task of training Generative Adversarial Networks (GANs) on a substantial scale while

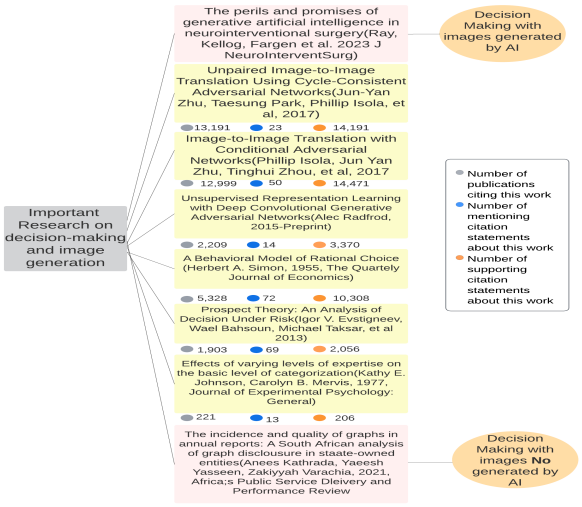

Figure 2: Significant Research Literature Impacted by Generative AI and Images and their Influence in Decision Making

specifically investigating the unique instabilities that arise within this expansive context. A noteworthy approach applied in this research involves the implementation of orthogonal regularization within the generator component of the GAN architecture. This strategic regularization technique is pivotal in stabilizing the training process and enhancing the quality and fidelity of the images generated. Consequently, Brock's study not only advances the capabilities of GAN-based image synthesis but also underscores the critical role of orthogonal regularization in addressing challenges associated with large-scale GAN training (Brock, Donahue, and Simonyan 2018).

The progress made in image generation has paved the way for diverse applications, encompassing style transfer, object transfiguration, season conversion, and photo enhancement (Zhu et al. 2017). The capacity emerges through acquiring a learned mapping between distinct image domains to engender images that embody the attributes intrinsic to specific styles or domains. For instance, the GANs method can imbue natural photographs with distinctive artistic characteristics to a collection of renowned painters, thereby transforming the images into their respective styles.

In the realm of decision-making concerning image generation, the onus predominantly rests upon users or stakeholders who employ the generated images for multifarious objectives. These decision-makers encompass individuals, entities, or even automated systems that depend on the synthesized images to facilitate subsequent analyses or decision-making protocols. The images generated are of utility across diverse domains, including arts, design, advertising, and scientific investigation. In the context of art and design, decision-makers can be artists, designers, or creative directors who use the generated images for inspiration or as integral components of their creative workflows.

Within scientific research, the role of decision-makers extends to researchers and scientists who harness the generated images to advance data analysis, facilitate visualization, and fortify their research findings. In this scientific context, these decision makers often depend on the images synthesized to derive conclusions and glean insights from the underlying data, thus enhancing the depth and scope of their investigative efforts. However, a crucial consideration emerges, underscoring the imperative of decision makers to demonstrate prudence and judicious reasoning when integrating generated images (Ribeiro, Singh, and Guestrin 2016). The significance of comprehending the rationales underpinning predictions is important in evaluating the trustworthiness of a model's outputs. This principle finds applicability within the realm of image generation, necessitating decision-makers to possess a lucid comprehension of the procedural intricacies governing image synthesis, as well as a discerning awareness of the inherent constraints within the algorithms driving the process

Figure 2 shows how research on creating images from generative AI techniques, such as GANs, involves training models to generate new images based on patterns and characteristics learned from training data (Zhu et al. 2017). However, there is sensitive information that AI can manipulate in poor democracies, where the quality of graphs can be misinterpreted. For instance, in South Africa, 64% of South African state-owned entities (SOEs) present graphs in their annual reports, with non-financial graphs being disclosed more than financial graphs. Using the graph discrepancy index (GDI), it was found that SOEs tend to overstate data trends more than understate trends, resulting in a better image of the SOE being presented (Kathrada, Yasseen, and Varachia 2021).

Reflection Questions Derived from Figure 2

Can generative AI prolong the power in poor democracies where information can be manipulated? If we know that a picture means two thousand words, how can we trust generative AI when it can play a role in critical decisions like budget spending on state-owned entities?

Societal Assessment in Neuro-interventional Surgery

The impact of generative AI in neuro-interventional surgery is a critical area of assessment, as it presents both opportunities and challenges in the realm of surgical practice. While the literature primarily focuses on the broader implications of AI in surgery, there is a growing recognition of the potential for AI, particularly generative AI, to support and enhance neuro-interventional surgical procedures.

Societal Assessment in Sensitive Information Using Graphs

Generative AI in state-owned entities (SOEs) prompts a comprehensive societal assessment that analyzes its ethical, legal, and technical implications. Ethically, generative AI's use in SOEs generates concerns about potential bias or misinformation, especially in financial reporting, reputation management, and crisis response strategies(Wu et al.

2016). Legally, employing generative AI in SOEs requires scrutiny to comply with disclosure requirements, perform materiality assessments in sustainability reporting, and adhere to corporate governance standards. Technically, evaluating generative AI in SOEs includes checking the accuracy, reliability, and transparency of AI-generated graphs and data, particularly in investment efficiency, green innovation, and emissions trading compliance contexts (Wang, Zhang et al. 2022). Therefore, generative AI's societal assessment in SOEs involves a thorough review of ethical, legal, and technical aspects to minimize potential risks and ensure the responsible, ethical use of generative AI in managing sensitive information and creating graphs within these entities.

Assessing Video Generation Challenges in Decision-Making Process

Video generation can exert a substantial influence on decision-making paradigms. Integrating AI systems within decision-making frameworks necessitates meticulously examining multifaceted aspects encompassing system design parameters, human-technology interface dynamics, and ethical considerations (Duan, Edwards, and Dwivedi 2019). Integrating AI into decision-making processes, encompassing scenarios involving videos holds the promise of elevating decision-making proficiencies. AI systems can adopt roles ranging from support and augmentation to replacement or even automation of decision-making. In response, scholars have advanced the development of design criteria and conceptual frameworks (Klein 2008).

The discourse surrounding the involvement of AI systems, including expert systems, in decision-making is rooted in analyzing decision structures, classified as structured, semi-structured, and unstructured decisions(Duan, Edwards, and Dwivedi 2019). Integration of AI within decision-making processes accentuates inquiries about equitable treatment, considerations of respect, trustworthiness, potential dehumanization, and the suitability of the decision-maker's role (Bankins et al. 2022). A comparative evaluation reveals that the utilization of human decision-makers often engenders more favorable perceptions concerning respectful treatment relative to AI counterparts (Bankins et al. 2022).

A pertinent framework in this context is the Evidence-Based Support for Decision-Making Practice Framework, designed to offer comprehensive guidance to facilitate decision-making for individuals with cognitive disabilities (Douglas and Bigby 2020). This framework places a notable emphasis on the comprehensive comprehension of the desires and preferences of these individuals and orchestrates the active participation of an array of stakeholders, encompassing families, support personnel, guardians, and healthcare practitioners (Douglas and Bigby 2020).

Within the healthcare landscape, an exploration is undertaken into applying generative AI for clinical diagnosis(Kulkarni and Singh 2023). The integration of AI has the potential to contribute to clinical diagnostic processes; however, a series of challenges emerge, necessitating a concerted effort to ensure the precision and reliability of clinical data utilized by AI systems (Kulkarni and Singh 2023).

Figure 3 shows how influential AI can be in promotional videos, while in medical care in 2009, an example was an intended effort to automate video images of patients ending their lives.

Reflection Questions Derived from Figure 3

Can generative AI exacerbate the influence of marketing unnecessary things in human minds? Can generative AI systems save radiologists' time without endangering patients?

Societal Assessment on Promotional Videos

Generative AI impacts societal assessment in the context of promotional videos through a multifaceted analysis that includes ethical, legal, and technical considerations. In terms of ethics, concerns arise about the potential for creating misleading or biased content in promotional videos using generative AI, especially when it comes to persuasive visual storytelling (Liu et al. 2019). it is crucial to evaluate the ethical implications thoroughly to ensure the content is transparent, fair, and free from manipulative elements. Legally, Kindylidi Cabral (2021) emphasize the need to scrutinize the deployment of generative AI in promotional videos to comply with regulations regarding deceptive advertising, intellectual property rights, and consumer protection laws. Technically, Dignum (2022) suggests evaluating the accuracy, reliability, and transparency of AI-generated promotional videos to align them with ethical and legal standards. Thus, generative AI's impact on societal assessment in promotional videos requires a comprehensive examination of ethical, legal, and technical dimensions to mitigate potential risks and ensure the responsible and ethical deployment of generative AI in creating promotional content.

Societal Assessment on Video Images at the End of Life

The assessment of the societal impact of generative AI using videos to support end-of-life care in hospitals encompasses a comprehensive evaluation of ethical, legal, and technical dimensions. Ethically, the use of generative AI in creating videos for end-of-life care necessitates careful consideration to ensure that the content is respectful, empathetic, and aligned with the values and preferences of patients and their families(Bloomer et al. 2022; Howes 2015). Moreover, the legal dimension involves scrutiny to ensure compliance with patient privacy regulations, informed consent, and the ethical use of patient data in creating supportive videos for end-of-life care in hospital settings(Dunphy et al. 2016; Kryworuchko et al. 2013). Additionally, the technical assessment of generative AI in this context involves evaluating the accuracy, appropriateness, and effectiveness of the videos in addressing the emotional and informational needs of patients and families facing end-of-life care, while also considering the potential impact on the quality of care and patient outcomes (Maetens et al. 2019; Ananth et al. 2023). Therefore, the societal assessment impacted by generative AI using videos to support end-of-life care in hospitals necessitates a thorough examination of ethical, legal, and technical

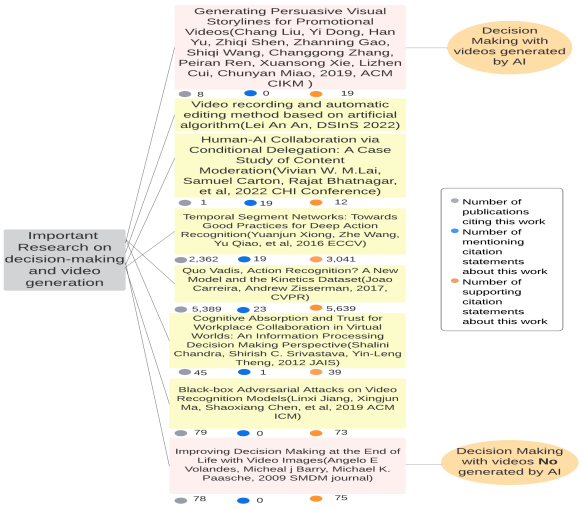

Figure 3: Significant Research Literature Impacted by Generative AI and Videos and their Influence in Decision Making

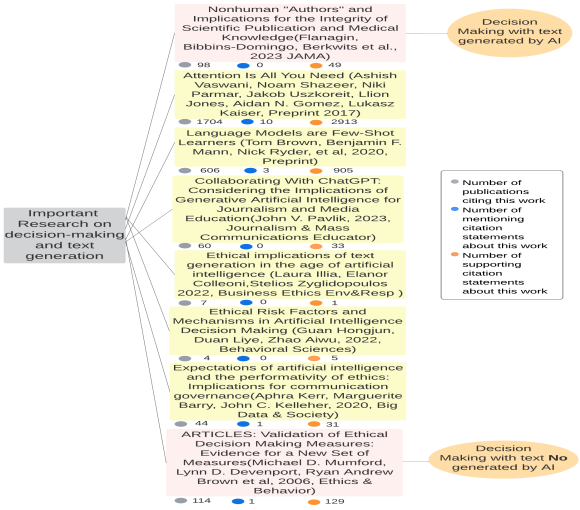

Figure 4: Significant Research Literature Impacted by Generative AI and Text and their Influence in Decision Making

dimensions to ensure that the use of AI technology in this sensitive context is respectful, supportive, and aligned with the best interests of patients and their families.

Assessing Text Generation Challenges in Decision-Making Process

Decision making is a complex process that involves evaluating different options and selecting the best course of action based on specific criteria or objectives. In recent years, generative AI text generation models have shown promise in helping with decision-making tasks. One notable model is the Transformer, a network architecture that relies solely on attention mechanisms, eliminating the need for recurrent or convolutional neural networks (Vaswani et al. 2017). This architecture is highly effective in sequence transduction tasks, such as machine translation. The Transformer model achieves state-of-the-art results in translation tasks, outperforming existing models and ensembles (Vaswani et al. 2017).

Integrating artificial intelligence (AI) into decision-making has raised many ethical considerations (Villegas-Galaviz and Martin 2023). The advent of AI can potentially widen the ethical gap, given that decisions are now encompassed within an enigmatic process that might not be entirely comprehensible to human agents (Villegas-Galaviz and Martin 2023). The ethics of care approach considers various factors, including situational dynamics, contextual nuances, interdependence, and susceptibility to influence, all playing crucial roles in AI-mediated decisions.

Figure 4 shows how text-based decision-making generated by AI can affect research and evidence support in the medical field. On the contrary, we can see that in 2016, there was rigorous research on the evaluation of a new set of measures by graduate students in the biological, health, and social sciences where AI was not involved.

Reflection Questions Derived from Figure 4

Large language models increasingly reply to prompts with a believably human response. Can they also mimic human behavior? If generative AI promises the automation of repetitive tasks and value creation in new jobs, how do data labelers, depending on the worker's location and the task assigned, earn from $1 per hour in Kenya to $25 per hour or more in the US?

Societal Assessment on Nonhuman "Authors" in Research

The integration of artificial intelligence (AI) in scientific writing has raised significant ethical and practical considerations. While AI presents opportunities for enhancing the efficiency and productivity of scientific manuscript preparation, it also introduces challenges related to transparency, accountability, and ethical dilemmas(Holmes et al. 2021). The use of AI, such as ChatGPT, in medical and scientific writing has prompted discussions on the ethical implications of AI-generated content and its potential to replace human authors(Holmes et al. 2021). Furthermore, the ethical considerations of AI in various domains, such as education, radiology, and agriculture, have been extensively examined, emphasizing the need for a community-wide ethical framework and the assessment of social and ethical impacts.

The multidisciplinary perspectives on AI underscore the emerging challenges, opportunities, and the need for a comprehensive research agenda in the context of AI integration in scientific writing and research. Additionally, the examination of epistemic modality in AI research articles highlights the evolving scholarly attitudes toward AI statements and their impact on scientific discourse. The societal and ethical implications of AI in agriculture, radiology, and surgery further emphasize the importance of understanding and addressing the broader societal impact of AI integration in scientific domains.

AI's Social Impacts and Individual Well-being: Methodological Challenges

The social impacts of artificial intelligence (AI) on individual well-being present methodological challenges that require careful consideration. The emergence of AI has led to its widespread impact on various sectors, necessitating an assessment of its effects on sustainable development (Vinuesa et al. 2020). The Institute of Electrical and Electronics Engineers (IEEE) has introduced a new standard, IEEE 7010, which focuses on assessing the well-being implications of AI, emphasizing the need to address the social and ethical implications of AI on human well-being (Schiff et al. 2020). Furthermore, the application of AI has brought about efficiency improvements and cost reductions, which are beneficial for economic growth, social development, and human well-being. However, it is crucial to recognize that certain AI applications, particularly appearance-focused activities on social media, can hurt vulnerable individuals' well-being, such as adolescents or those with low self-esteem (Javornik et al. 2022). Additionally, the widespread use of AI technology raises privacy and social justice issues, highlighting the need to address these concerns to safeguard individual well-being. Therefore, addressing the social impacts of AI on individual well-being requires a comprehensive approach that considers the ethical, social, and economic implications to ensure the responsible development and deployment of AI systems.

The impact of AI-based decision-making systems on individuals and societies has been a subject of concern, particularly about gender bias. Studies suggest that AI-based decision-making systems may be biased towards gender, thereby impacting individuals and societies. Additionally, the impact of AI on decision-making extends to the doctor-patient relationship, with concerns raised about the potential decrease in doctors' control over decision-making and the threat to professional autonomy posed by AI. Moreover, the impact of AI on decision-making has implications for community well-being, as community well-being indicators can be used to understand the impacts of AI on community well-being and aid in decision-making.

Conclusions

The advent of generative AI, exemplified by tools like ChatGPT, DALL-E, and GitHub Copilot, has sparked a new era of content creation and interaction. The simplicity and user-friendliness of these tools are evident, but so are the ethical complexities underlying their use. The ethical implications of AI decision-making extend to various domains. In image generation, algorithmic bias, and data misuse require careful attention. Video generation introduces considerations of equitable treatment, respect, trustworthiness, and the potential dehumanization of decision-making processes. In text generation, the ethics of care framework emphasizes situational dynamics, contextual nuances, and vulnerability in AI-mediated decisions.

The integration of generative AI into decision-making raises critical questions. Can AI exacerbate the influence of marketing unnecessary products and services? Can it streamline complex medical diagnostics without compromising patient safety? How might AI's role in decision-making impact poor democracies, sensitive information, or the credibility of crucial data-driven choices? These inquiries underscore the delicate balance between leveraging AI's potential and ensuring that ethical, legal, and societal considerations are meticulously addressed.

As AI's capabilities expand, responsible and transparent development practices must be at the forefront of AI research and deployment. Ethical frameworks, legal principles, and proactive measures to mitigate bias, enhance transparency, and foster accountability are essential to ensuring that integrating AI-powered generative algorithms in decision-making remains a force for positive change rather than a source of unintended consequences. By addressing these questions and challenges head-on, we can navigate the evolving landscape of AI-driven decision-making with a commitment to societal well-being and ethical integrity.

Acknowledgments

The authors acknowledge NIH BioMed Grant Number 150108136 and CMC Grant Number 12-09-C13057 under Florida A&M University.

References

Ananth, P.; Lindsay, M.; Mun, S.; McCollum, S.; Shabanova, V.; de Oliveira, S.; Pitafi, S.; Kirch, R.; Ma, X.; Gross, C. P.; et al. 2023. Parent Priorities in End-of-Life Care for Children With Cancer. *JAMA Network Open*, 6(5): e2313503–e2313503.

Bankins, S.; Formosa, P.; Griep, Y.; and Richards, D. 2022. AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, 24(3): 857–875.

Basit, A.; Zafar, M.; Liu, X.; Javed, A. R.; Jalil, Z.; and Kifayat, K. 2021. A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems*, 76: 139–154.

Bloomer, M. J.; Poon, P.; Runacres, F.; and Hutchinson, A. M. 2022. Facilitating family needs and support at the end of life in hospital: A descriptive study. *Palliative Medicine*, 36(3): 549–554.

Bohannon, J. 2015. Fears of an AI pioneer.

Brock, A.; Donahue, J.; and Simonyan, K. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

Bullock, J. B. 2019. Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7): 751–761.

Crawford, K. 2013. The hidden biases in big data. Harvard Business Review, April 1.

Davenport, T.; and Kalakota, R. 2019. The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2): 94.

Douglas, J.; and Bigby, C. 2020. Development of an evidence-based practice framework to guide decision making support for people with cognitive impairment due to acquired brain injury or intellectual disability. *Disability and rehabilitation*, 42(3): 434–441.

Duan, Y.; Edwards, J. S.; and Dwivedi, Y. K. 2019. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International journal of information management*, 48: 63–71.

Dunphy, E. J.; Conlon, S. C.; O'Brien, S. A.; Loughrey, E.; and O'Shea, B. J. 2016. End-of-life planning with frail patients attending general practice: an exploratory prospective cross-sectional study. *British Journal of General Practice*, 66(650): e661–e666.

Elahi, B. 2020. Is AI in Healthcare Doomed, or Destined for Greatness? *Journal of System Safety*, 55(3): 23–27.

Eppler, M.; Chu, T.; Gill, I.; and Cacciamani, G. 2023. The benefits and dangers of artificial intelligence in healthcare research writing. *Uro-Technology Journal*, 7(1): 01–02.

Esmaeilzadeh, P. 2020. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC medical informatics and decision making*, 20(1): 1–19.

Griffiths, F. T. 1996. " Sorcery is Dialectical": Plato and Jean Toomer in Charles Johnson's The Sorcerer's Apprentice. *African American Review*, 30(4): 527–538.

Guan, H.; Dong, L.; and Zhao, A. 2022. Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making. *Behavioral Sciences*, 12(9): 343.

Halaweh, M. 2018. Artificial intelligence government (Gov. 3.0): the UAE leading model. *Journal of Artificial Intelligence Research*, 62: 269–272.

Henman, P. 2019. ASSESSING ETHICAL AI-BASED DECISION-MAKING: TOWARDS AN APPLIED ANALYTICAL FRAMEWORK. *AoIR Selected Papers of Internet Research*.

Hole, K. J. 2023. Tools with general AI and no existential risk. *AI and Ethics*, 1–8.

Holmes, W.; Porayska-Pomsta, K.; Holstein, K.; Sutherland, E.; Baker, T.; Shum, S. B.; Santos, O. C.; Rodrigo, M. T.; Cukurova, M.; Bittencourt, I. I.; et al. 2021. Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 1–23.

Howes, J. 2015. Nurses' perceptions of medication use at the end of life in an acute care setting. *Journal of Hospice & Palliative Nursing*, 17(6): 508–516.

Javornik, A.; Marder, B.; Barhorst, J. B.; McLean, G.; Rogers, Y.; Marshall, P.; and Warlop, L. 2022. 'What lies behind the filter?'Uncovering the motivations for using augmented reality (AR) face filters on social media and their effect on well-being. *Computers in Human Behavior*, 128: 107126.

Jebari, K.; and Lundborg, J. 2021. Artificial superintelligence and its limits: why AlphaZero cannot become a general agent. *AI & SOCIETY*, 36(3): 807–815.

Jiang, Z.; Xu, F. F.; Araki, J.; and Neubig, G. 2020. How can we know what language models know? *Transactions of the Association for Computational Linguistics*, 8: 423–438.

Kathrada, A.; Yasseen, Y.; and Varachia, Z. 2021. The incidence and quality of graphs in annual reports: A South African analysis of graph disclosure in state-owned entities. *Africa's Public Service Delivery & Performance Review*, 9(1): 11.

Klein, G. 2008. Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50: 456–460.

Kryworuchko, J.; Hill, E.; Murray, M. A.; Stacey, D.; and Fergusson, D. A. 2013. Interventions for shared decision-making about life support in the intensive care unit: A systematic review. *Worldviews on Evidence-Based Nursing*, 10(1): 3–16.

Kulkarni, P. A.; and Singh, H. 2023. Artificial Intelligence In Clinical Diagnosis. *Jama*, 330: 317.

Liu, C.; Dong, Y.; Yu, H.; Shen, Z.; Gao, Z.; Wang, P.; Zhang, C.; Ren, P.; Xie, X.; Cui, L.; et al. 2019. Generating persuasive visual storylines for promotional videos. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 901–910.

Lysaght, T.; Lim, H. Y.; Xafis, V.; and Ngiam, K. Y. 2019. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. *Asian Bioethics Review*, 11: 299–314.

Maetens, A.; Beernaert, K.; De Schreye, R.; Faes, K.; Annemans, L.; Pardon, K.; Deliens, L.; and Cohen, J. 2019. Impact of palliative home care support on the quality and costs of care at the end of life: a population-level matched cohort study. *BMJ open*, 9(1): e025180.

Ntoutsi, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdl, W.; Vidal, M.-E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. 2020. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3): e1356.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144.

Robbins, S. 2020. AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & SOCIETY*, 35: 391–400.

Schiff, D.; Ayesh, A.; Musikanski, L.; and Havens, J. C. 2020. Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, 2746–2753. Ieee.

Schmidt, T. 2018. Solving the AI Race: Addressing the potential pitfalls of competition towards artificial general intelligence.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Villegas-Galaviz, C.; and Martin, K. 2023. Moral distance, AI, and the ethics of care. *AI & society*, 1–12.

Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S. D.; Tegmark, M.; and Fuso Nerini, F. 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, 11(1): 1–10.

Vold, K.; and Harris, D. R. 2021. How Does Artificial Intelligence Pose an Existential Risk?

Wang, W.; Zhang, Q.; et al. 2022. Does Green Finance Reform Promote Corporate Green Innovation? Evidence from a Quasi-Natural Experiment. *Mathematical Problems in Engineering*, 2022.

Wiener, N. 2019. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press.

Wu, P.; Gao, L.; Chen, Z.; and Li, X. 2016. Managing reputation loss in China: in-depth analyses of financial restatements. *Chinese Management Studies*, 10(2): 312–345.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.