Timothée de Raucourt
Undergraduate Student, Software Engineering
York University
1234 University Ave, Toronto, ON
(123) 456 7890
timothee@my.yorku.ca

Nima Talebzadeh Gargari and Milad Shakeri Bonab
York University
4700 Keele St, North York
Toronto, ON, M3J 1P3

Dear Professors Talebzadeh Gargari and Shakeri Bonab,

I hope this letter finds you well. Attached is the report titled "*Addressing AI's Impact on Human Autonomy, Decision-Making, and Ethical Development.*" This report was prepared following our agreement dated July 26th, 2024, and it explores several critical issues surrounding artificial intelligence.

The report addresses three main challenges articulated by Zittrain: the reduction of human autonomy due to AI, the potential for AI to enhance human reflection and decision-making, and the need for ethical evolution in AI systems. Its purpose is to define and clarify the relationship between AI and human autonomy within the context of software engineering, and to propose a technological solution involving ethical development to these challenges. Please note that the report does not delve into deep technical algorithmic details, cover every possible application of AI, or attempt to predict the future trajectory of AI technology.

Please review the enclosed report at your convenience. If you have any questions or require further information, please feel free to contact me. I look forward to your feedback and am available for any necessary discussions or clarifications.

Thank you for your time and consideration.

Sincerely,
Timothée de Raucourt
Undergraduate Student, Software Engineering
York University
1234 University Ave, Toronto, ON
(123) 456 7890
timothee@my.yorku.ca

Addressing AI's Impact on Human Autonomy, Decision-Making and Ethical Development

Prepared for

Professors
Nima Talebzadeh Gargari
Milad Shakeri Bonab

Toronto, Ontario
August 18, 2024

by
Timothée de Raucourt
Student, Software Engineering
York University
timothee@my.yorku.ca

Table of Contents

1.0 Introduction

The professor, Jonathan Zittrain of Computer Science at the Harvard Kennedy School of Government expresses concern in human autonomy reduction due to the recent rise of artificial intelligence implicated in our technological systems and infrastructures. Conversely, he proposes further investigation in helping human reflection and decision-making rather than using artificial intelligence as its replacement. He worries the "set it and forget it" mindset could lead to regret as the technology evolves and becomes more interdependent in our systems, leading to complexity in managing and maintaining its decision-making and ethical dimensions as time progresses.

This report addresses three critical challenges expressed by Zittrain: the reduction of human autonomy due to AI, the use of AI in enhancing human reflection and decision-making, and the necessity for ethical evolution in AI systems. The report intends to define, clarify and educate about artificial intelligence and its ethical implications with human autonomy and the discipline of software engineering, then proposing a technological solution. It will not delve into deep technical algorithmic details, cover every possible application of artificial intelligence, nor will it try to predict the future of artificial intelligence. The report begins with an introduction defining human autonomy and its importance, followed by the current state of artificial intelligence. Then it examines the effect and areas on human autonomy. Ethical dimensions are analyzed and discussed followed by how artificial intelligence can be used to enhance human reflection and decision-making and proposing solutions for ethical development of artificial intelligence. The report concludes with a summary of the crux and reinforcement of the central message.

2.0 Background

2.1 Definition of Artificial Intelligence

Artificial Intelligence (AI) simply refers to systems and technology that have the ability to emulate human thought processes and perform tasks that normally require human intelligence [7], [9]. Large volumes of data are processed by AI systems using models and algorithms, which allows them to predict outcomes, categorize information, and interact with their surroundings in ways that are similar to the way humans operate. These exercises cover a broad variety of skills, including learning, thinking, reasoning, problem-solving, perception, decision-making and language comprehension. It strives to imitate these cognitive processes so they can successfully interact with their surroundings and learn new information. AI covers a broad spectrum of technologies, such as robotics, computer vision, natural language processing, deep learning, and machine learning.

2.1.1 Artificial Intelligence vs Machine Learning

Despite the common confusion between the terms machine learning and AI, machine learning is a subset of AI. Machine learning refers to the techniques and technologies that enable systems to see patterns, make decisions, and enhance themselves through data. [9]. Through the study and application of machine learning, software engineers explore the boundaries of what can be done to enhance the capabilities of AI systems such as perception, logic, and behavior.

2.1.2 Artificial Intelligence vs Algorithms

The formal definition of an algorithm is "a process or set of rules to be followed in calculations or other problem-solving operations." A common misconception is to confuse AI, particularly machine learning, with traditional algorithms. Deep Blue is a chess-playing supercomputer being a prime example of this confusion. In the late 1990s, Deep Blue was able to defeat human chess champions, but it did not use AI or machine learning to achieve this. Fundamentally, however, it was based on tree search algorithms that analyzed millions of possible move sequences in a systematic way [11], [12]. Deep Blue was able to determine the optimal results by using these algorithms, which represented every move that might be made at every level of the game. The system was developed in collaboration with several chess grandmasters, who provided their expertise to translate human chess strategies into tens of thousands of explicit rules for playing at a grandmaster level [7]. Thus, Deep Blue operated based on predefined rules and computational power, not through learning or adapting like modern AI systems.

2.2 Mechanisms of Artificial Intelligence

The mechanisms of AI are fundamental to understanding its role and impact on human autonomy, decision-making, and ethical considerations. Machine learning algorithms can be divided into four types according to distinct characteristics of the training data [7].

Supervised Learning is a method by which human supervisors provide prelabeled input and output data. The algorithm must find a method to reach those inputs and outputs by making observations and predictions by being manually corrected by the human supervisor until the algorithm reaches high accuracy or performance [13].

Unsupervised Learning allows the machine learning algorithm to study and interpret unlabeled data on its own. This method tries to organize data by finding similarities, differences and patterns in the data set and may discover hidden patterns or structures without the need of human intervention or bias [14].

Semi-Supervised Learning combines labeled and unlabeled data, balancing human input and algorithmic pattern recognition. This combines the pros and cons of supervised and unsupervised learning and allows the machine to learn to label unlabeled data [7].

Reinforcement Learning uses agents provided with a set of rules, actions, parameters and end values. The agents interact with their environment with trial and error and get rewarded and punished accordingly. Rewarding behavior signals to the agent to repeat the same strategies the following time in order to progress closer to the end values [15].

Regardless of the model type, one of the biggest challenges in machine learning is aggregating high quality training data. The better the data represents the problem, the better the results. Having poor quality data or insufficient data can lead to decreased accuracy and trust followed by ethical issues down the line depending on the applications of the system.

2.3 Definition of Human Autonomy

The essence of human autonomy is the capacity of self-ruling, self-government and self-determination. It is related to the practical reasoning that people possess the ability to evaluate the justifications for taking actions, to seek goals that are deemed valuable, and the ability to resist irrational impulses [3]. Autonomy is about having authentic choices and is a vital component of human dignity and agency. As an ethical value, human autonomy encompasses the essential principle of freedom: authoring your own life, choosing alternatives, making important decisions, developing a sense of desires and values, without unintended interference of others' guidance or influence. Respecting autonomy is important because it is valuable both as a standalone value and as a component of human dignity and well-being.

2.4 Role of Software Engineering in Ethical AI Development

The methodical application of engineering principles to the design, development, testing, and maintenance of software systems is known as software engineering. This pertains to AI and entails developing scalable and modular structures that enable AI systems to handle massive amounts of data and smoothly interface with other software elements. Engineers create data pipelines and choose the right technology to guarantee AI models are capable of handling practical applications.

But there are difficulties in the creation of AI systems, especially in the area of ethics. Data is a major component of AI systems, and creating representative and high-quality data sets is crucial to creating accurate models. Ethical issues with AI systems, like prejudice, privacy, and accountability, must be addressed. Engineers need to put policies in place to guarantee openness, lessen bias, safeguard user data, and foster confidence in AI systems. In order to guarantee that AI systems be utilized responsibly and do not reinforce negative prejudices or infringe privacy, several ethical issues are essential.

3.0 Transformative Impact of AI in Society and Infrastructure

3.1 Recent Rise and Development of AI

Over the past few years, AI has experienced a considerable increase in both development and incorporation into numerous aspects of modern life. This increase is the result of developments in machine learning, immense collection of data, breakthroughs in mathematical modeling, and computing power, which have made it possible for AI systems to carry out jobs that were previously believed to be the responsibility of humans. AI has advanced quickly, from natural language processing to computer vision, and is now used in a variety of businesses outside of research labs. The way we live and work has been profoundly impacted by the broad adoption of AI technologies in industries including healthcare, banking, entertainment, and transportation. AI research is trending away from technology-oriented applications that prioritize performance and productivity and toward applications that prioritize humanity, which emphasize combining artificial intelligence with human intelligence to enhance human intelligence [4]. As seen in Figure 1, there has been a recent rise in research papers focused on collaboration between humans and AI for augmenting human decision and interactions for tasks.
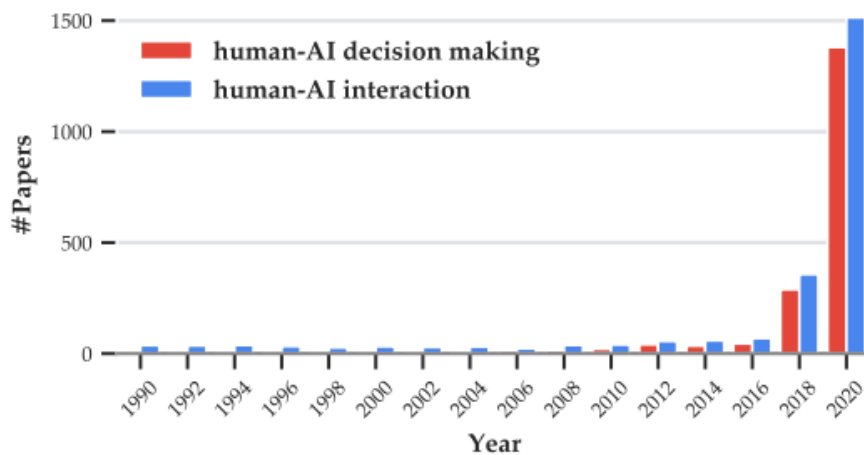
Fig. 1. The amount of publications produced over the years on Google Scholar for the two keywords, human-AI interaction and human-AI decision making [5].

3.2 The Increasing Importance of Understanding AI

AI is becoming more and more integrated into daily life, which means that people, businesses, and society as a whole must understand this technology. Understanding AI's possible effects, limitations, and capabilities is becoming important as it becomes more integrated into everyday activities and decision-making processes. This knowledge is crucial for both maximizing the advantages of AI and reducing its drawbacks, including bias in AI models, technological limitations, privacy invasion, and ethical dilemmas. In order to ensure that users are knowledgeable and equipped to negotiate the challenging terrain that AI provides, there is an increasing demand for a thorough and easily accessible knowledge base about AI technology.

3.3 Applications of AI in People and Society

AI has become a vital component of everyday life, revolutionizing communication, interaction, and decision-making. Without people even recognizing it, AI is being incorporated into both personal and societal practices.

Virtual assistants such as Alexa, Siri and Google Assistant, support us in our homes by responding to voice requests and doing a variety of activities, such as managing smart home devices, playing music and creating reminders.

Chatbots based on Large Language Models such as ChatGPT and Google Gemeni are especially important because they offer advanced conversational capabilities and help with a variety of activities, such as answering difficult queries and drafting emails. Large volumes of data are used by these models to produce responses that are human-like and to enable interesting, educational interactions.

Recommendation AI systems incorporated by social media with websites like Facebook, Instagram, and Twitter to analyze user behavior patterns and recommend material that matches user tastes, so personalizing the user experience.

AI-powered data collection and analysis analyzes behavior, preferences, and trends through online activity monitoring. This results in targeted advertising based on search queries, browsing history, and social media engagement, often without the user's knowledge or consent.

Although these applications improve user interaction, they also bring up moral issues like the development of echo chambers, cognitive offloading, swaying public opinion, and invasions of privacy.

3.4 Applications of AI in High-State Domains and Infrastructure

There are dependent and interdependent reliance on AI within our societal infrastructure that help decision-making within our society. Table 1, which was modified from [5] to better fit the report's emphasis on ethical issues with AI, classifies the different kinds of decision tasks into application categories for human-AI decision-making in high-stake domains. The paper reveals the current complements of AI and human decision making to assist in deciding tasks and uncovers domains that could cause dangerous repercussions such as total automation with AI being undesirable in such sectors.

Table 1. Types of decision tasks grouped by application domains in human-AI decision making [5]

Domain	Decision task
Law & Civic	Recidivism prediction and its slight variations, likelihood to recidivate, bail outcomes prediction, child maltreatment risk prediction
Medicine & Healthcare	Medical disease diagnosis, cancer image search, cancer image classification, COVID-19 diagnosis, balance disorder diagnosis, clinical notes annotation/medical coding, stroke rehabilitation assessment
Finance & Business	Income prediction, loan approval, loan risk prediction, sales forecast, property price prediction, apartment price prediction, selecting overbooked airline passengers for re-routing, determining to freeze bank accounts due to money laundering suspicion, stock price prediction, marketing email prediction, dynamically pricing car insurance premiums
Education	Students' performance forecasting, student admission prediction, student dropout prediction, LSAT question answering
Leisure	Music recommendation, movie recommendation, song rank order prediction, speed dating, Facebook news feed prioritization, Quizbowl, draw-and-guess word guessing, chess playing, plant classification, goods division
Professional	Job promotion, meeting scheduling assistance, email topic classification, cybersecurity monitoring, profession prediction, military planning (monitor and direct unmanned vehicles)
Others	Deception detection, forest cover prediction, toxicity classification, nutrition prediction, person weight estimation, attractiveness estimation, activity recognition, emotion analysis, religion prediction

With the use of AI and humans for decision-making processes, we see the potential use for providing predictions and recommendations for decisions. Such as in medicine and healthcare, guidance with AI alleviates inherent challenges in comprehending and predicting biology-related processes, therefore providing insight and allowing practitioners to follow their own decision. We see how AI is embedded into societal infrastructure with select examples such as in finance by dynamically pricing car insurance premiums and in the professional domain with hiring, firing and promotions in the workplace. While using AI in collaboration with humans to assist decision-making, total automation in certain domains raises ethical concerns with hidden bias, justice, reasoning and their accountability for its decisions.

4.0 Effects of AI on Human Autonomy

Human autonomy is a critical factor to take into account in the context of AI as these systems have a growing impact on human interactions and decision-making processes. As AI is incorporated into more fields, concerns are raised regarding how much this technology may increase or decrease personal freedom. By their very nature, AI systems are meant to supplement human talents. They may help with

everything from complicated decision-making situations to everyday jobs. A rising number of people are worried that an over-reliance on AI may compromise human autonomy and lessen a person's ability to make their own judgements.

4.1 Benefits of AI on Human Autonomy

By enhancing capacities, lessening cognitive load, and promoting physical functioning, AI improves human autonomy. When AI systems are embedded into physical environment interactive machines, they can shield people from harm and facilitate their movement. For example, autonomous cars can help individuals with disabilities by enabling them to drive themselves without the assistance of human drivers. In the same way, AI-powered medical devices allow people to manage their health by monitoring conditions and making timely interventions.

Furthermore, AI respects users' autonomy by providing tailored recommendations. These suggestions, in contrast to decisions, are considerate of individual tastes and can help people make well-informed choices. Personalized AI systems, for instance, make product recommendations on e-commerce platforms based on user behavior in the past, improving the user experience without forcing the user to make judgements.

A significant advantage of AI is its ability to offload cognitive tasks to external devices like smartphones. Large volumes of information can be stored and accessed by users, allowing them to review what they know whenever necessary. By using current insights rather than out-of-date material, this offloading enables people to operate as knowledge experts. Professionals can improve their decision-making and strategic planning, for instance, by using AI to access real-time data.

Furthermore, by dispersing tasks throughout an individual's environment, AI-powered cognitive offloading conserves people's own cognitive resources. Offloading can enhance immediate task performance by speeding it up and lowering errors, according to studies. It also improves the way unrelated future activities and secondary tasks are performed. AI-driven navigation systems, for example, save people from the burden of learning routes and enable them to concentrate on driving safely, which can enhance task performance in general.

4.2 Detriments of AI on Human Autonomy

By taking advantage of flaws in human decision-making, AI programs can exert influences on users. They possess the ability to gently alter and affect human decisions, one of the major obstacles to sustaining human autonomy in the era of AI [3]. AI-based targeted ads, for example, have the ability to trick consumers into making purchases they otherwise might not have thought about. These manipulations serve the manipulator's objectives rather than promoting true autonomy since they cause erroneous beliefs and obstruct the user's ability to reason. This gives rise to worries about the loss of autonomy since people can become dependent on recommendations from AI, which would impair their ability to think critically and independently.

Although cognitive offloading offers advantages, it also has significant disadvantages. When tasks and related information are complicated or challenging, cognitive offloading rises. On the other hand, research has indicated that an over-reliance on offloading may result in a decline in the memory performance of the offloaded information [16]. For example, utilizing navigation systems for spatial tasks might lead to spatial memory impairments, making it harder to remember routes on your own. This suggests that offloading enhances task performance in the short term but may impede the creation of long-term memories, hence decreasing cognitive autonomy.

5.0 Enhancing Human Reflection and Decision-Making with AI

As suggested by Professor Zitrain, further investigation should be conducted in helping human reflection and decision-making rather than using artificial intelligence as its replacement. Properly executed integration of AI across different domains can help us promote more knowledgeable, effective, and autonomous choices.

5.1 Chatbots in Education

With its ability to provide interactive learning, automated essay grading, and individualized tutoring, chatbots such as ChatGPT and Google Gemini have potential in the education domain [8]. These systems provide personalized instructional support that may adjust to each user's unique learning needs and preferences by utilizing natural language processing techniques. AI chatbots can evaluate a student's areas of strength and weakness and provide tailored resources and feedback to improve learning outcomes in individualized tutoring. Essay grading that is automated can assess written assignments consistently and effectively, giving teachers thorough feedback and sparing them time. Additionally, language barriers are eliminated by AI-powered language translation services, creating more inclusive learning environments. While adaptive learning systems continuously modify the difficulty of materials based on student performance, interactive learning platforms powered by AI may engage students with dynamic content and simulations. Because AI can summarize materials and create test questions, it can assess and comprehend important topics more quickly, which improves the quality of the educational process [4].

5.2 Predictive AI Enhancing Decision-Making

As seen in Section 3.4, the predictive potential of AI are proven to be extremely useful in domains like professional management, finance, and health. AI systems are used in healthcare to evaluate enormous volumes of biological data and offer insights into complicated processes, helping practitioners make more precise diagnosis and treatment decisions. For instance, AI may identify possible health hazards, forecast disease outbreaks, and recommend individualized treatment plans. In the finance industry, AI algorithms use a variety of data, such as driving habits and risk assessments, to dynamically determine auto insurance rates. Both insurers and policyholders profit from this strategy's ability to provide more accurate and customized pricing. AI is also having a big impact on the workplace, helping with recruiting, firing, and promotion processes. AI can make suggestions that assist impartial and fair decision-making processes by examining market trends and employee performance data.

6.0 Ethical Concerns with AI

6.1 Bias

The possibility of bias in algorithms is one of the most important ethical problems with AI. If AI systems are educated on biased data sets, they may unintentionally reinforce or worsen societal biases. This may result in people being treated unfairly because of their socioeconomic situation, gender, or race [6]. In order to address prejudice, data must be carefully chosen and preprocessed. Continuous monitoring is also necessary to make sure AI systems do not perpetuate already-existing inequities.

6.2 Lack of Transparency

Because AI systems are frequently opaque and complex, it might be challenging for users to comprehend the decision-making process [6]. In order to foster confidence and guarantee that users can hold AI systems responsible, transparency is essential. This entails explaining AI-driven results in plain terms and making algorithms and decision-making procedures easier for non-experts to comprehend.

6.3 Acccountability of Decisions

Making judgments with AI systems and deciding who is responsible for them is another major challenge. It can be difficult to determine who is at fault when AI systems make bad or incorrect decisions such as developers, users, or the AI itself [6]. The implementation of well-defined accountability frameworks is important in order to effectively address the fallout from AI's actions and guarantee the existence of redress mechanisms.

6.4 Data Protection

Data privacy is a concern since using AI frequently entails gathering and analyzing vast volumes of personal information. It is imperative to guarantee that the collection, storage, and utilization of data adheres to the private rights of individuals. This entails giving consumers control over their personal data and putting strong data protection mechanisms in place [6].

7.0 Ensuring Ethical Evolution in AI Systems

7.1 Human-Centered Design for Autonomy

Using a human-centered design approach is an essential component of the ethical development of AI. Rather than replacing human autonomy, this method focuses on strengthening and promoting it. Rather than taking users' place when making decisions, AI systems should be built to empower users by offering resources and insights that facilitate decision-making. AI can, for instance, provide diagnostic recommendations in the medical field so that physicians can make well-informed decisions instead of being told what to do. AI may collaborate with humans in decision-making processes by placing a high value on human agency. This promotes a cooperative relationship between humans and technology.

7.2 Mitigating Bias in AI Systems

AI systems that are biased may produce unjust results that either reinforce already-existing disparities or create new ones. Software engineers are essential in locating, reducing, and eliminating bias from AI systems. This entails creating algorithms that are both technically and socially conscious. Bias can be reduced by employing strategies like varied training datasets, bias detection techniques, and fairness-aware machine learning. Furthermore, to guarantee that AI systems continue to be impartial and fair throughout time, continuous observation and assessment are required.

7.3 Transparency and Explainability

AI systems must give explicit explanations for their predictions and decisions in order to be ethically acceptable. Because of this transparency, users are able to understand and, if needed, challenge the outcomes generated by AI. Retaining faith in AI requires explainability, or the ability of systems to justify their outputs. Users are better able to evaluate an AI system's validity and relevance when they are aware of the process by which it arrived at a certain choice. This is especially crucial in high-stakes fields like criminal justice, healthcare, and finance where AI choices can have far-reaching effects.

7.4 Data Privacy and Security

Reliable data privacy and security measures are essential for safeguarding personal information in artificial intelligence systems. Data security is a crucial ethical concern as AI depends more and more on massive databases, many of which contain sensitive information. To stop unwanted access and data breaches, software engineers need to put strong encryption, anonymization, and access control methods in place. Additionally, openness in the use of data such as giving people explicit information about how their data will be used and safeguarded builds trust and guarantees adherence to moral and legal requirements.

8.0 Discussion: Addressing the Challenges

AI systems provide serious obstacles to human autonomy and moral decision-making as they grow more commonplace. Many AI systems are opaque and sophisticated, which might make it difficult to understand how decisions are made and potentially reduce people's agency and sense of control. In order to tackle these issues, a few crucial tactics are required.

First, by using a human-centered design methodology, AI systems are guaranteed to strengthen rather than diminish user autonomy. In order to meet their needs and ideals, this entails including users in the development process and creating inclusive and accessible systems.

In order to stop AI systems from maintaining social injustices, bias mitigation is essential. Software engineers may design more equitable systems that uphold human autonomy by utilizing a variety of data sets and putting in place mechanisms for identifying and removing biases. To ensure equity and deal with new prejudices, audits must be conducted on a regular basis.

Sustaining trust and autonomy requires explainability and transparency. Clear justifications for decisions made by AI systems should be given so that people may comprehend and question the results. Users are able to maintain control and make wise decisions because of this transparency.

It's also critical to establish accountability structures. AI systems are held to ethical norms and their negative effects are efficiently addressed when developers and users have well defined roles and responsibilities, as well as procedures for reporting problems and finding remedies.

Strong data privacy and security procedures are necessary to protect personal data. Users should be in charge of their data, with clear consent procedures, transparent data gathering methods, and the ability to access, amend, and delete their information.

9.0 Conclusion

The purpose of this report was to address three critical challenges expressed by Zittrain: the reduction of human autonomy due to AI, the use of AI in enhancing human reflection and decision-making, and the necessity for ethical evolution in AI systems.

While Zittrain expresses reduction in human autonomy, AI benefits human autonomy with the use of physically embedded AI systems that interact with their environment in a way that supports self-determination and self-ruling, supporting personalized content and lessening demanding cognitive tasks. However, human autonomy is negatively affected by AI altering and affecting human decisions, recommendation systems having data trained on outdated user data and by also allowing cognitive offloading.

Human reflection and decision-making can be enhanced in education with the use of chatbots by providing interactive learning, automated essay grading, and individualized tutoring. Other high-stake domains such as finance and healthcare benefit greatly with predictive models that support collaboration between human and AI for decision-making allowing humans to gain valuable insight while following their own decisions.

Ethical concerns such as bias, transparency, judgement and accountability are the primary concerns of AI in our society and infrastructure. Software engineers developing AI technology are proposed to incorporate human-centered design for human autonomy, place systems to detect and mitigate bias, open transparency and explainability and ensure data protection in AI systems.

Understanding the implications of AI and human autonomy is important for individuals realizing its influence, but also discovering how AI can be used to aid in bettering reflection and decision-making. As AI continues to be developed in the coming years, software engineers should ensure ethical development of its systems.

10.0 References

[1] Y. K. Dwivedi, A. Sharma, N. P. Rana, M. Giannakis, P. Goel, and V. Dutot, "Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions," Apr. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162523002640. [Accessed: Jul. 29, 2024].

[2] Á. A. Cabrera, M. T. Ribeiro, B. Lee, R. Deline, A. Perer, and S. M. Drucker, "What Did My AI Learn? How Data Scientists Make Sense of Model Behavior," Mar. 2023. [Online]. Available: https://dl.acm.org/doi/full/10.1145/3542921. [Accessed: Jul. 29, 2024].

[3] A. Laitinen and O. Sahlgren, "AI Systems and Respect for Human Autonomy," *Frontiers in Artificial Intelligence*, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2021.705164/full. [Accessed: Jul. 29, 2024].

[4] S. J. H. Yang, H. Ogata, T. Matsui, and N.-S. Chen, "Human-centered artificial intelligence in education: Seeing the invisible through the visible," *Journal of Artificial Intelligence in Education*, vol. 27, no. 2, pp. 215-234, 22 Jan. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666920X21000023 [Accessed: Jul. 29, 2024].

[5] V. Lai, C. Chen, Q. V. Liao, A. Smith-Renner, and C. Tan, "Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies," Dec. 2021. [Online]. Available: https://arxiv.org/abs/2112.11471. [Accessed: Jul. 29, 2024].

[6] A. Z. Huriye, "The Ethics of Artificial Intelligence: Examining the Ethical Considerations Surrounding the Development and Use of AI," *American Journal of Technology*, vol. 4, no. 2, pp. 45-60, Apr. 2023. [Online]. Available: https://gprjournals.org/journals/index.php/AJT/article/view/142 [Accessed: Aug. 6, 2024].

[7] G. Allen, "Understanding AI Technology," *Joint Artificial Intelligence Center* (JAIC), Department of Defense, Apr. 2020. [Online]. Available: https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf [Accessed: Aug. 6, 2024].

[8] D. Baidoo-Anu and L. O. Ansah, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning," *Journal of Artificial Intelligence*, vol. 7, 2023. [Online]. Available: https://dergipark.org.tr/en/pub/jai/issue/77844/1337500. [Accessed: Aug. 6, 2024].

[9] "Artificial Intelligence (AI) vs. Machine Learning," *Columbia University AI Engineering*, [Online]. Available: https://ai.engineering.columbia.edu/ai-vs-machine-learning/. [Accessed: Aug. 6, 2024].

[10] "How Do AI and Algorithms Differ From Each Other?," *SDLCCorp*, [Online]. Available: https://sdlccorp.com/post/how-do-ai-and-algorithms-different-from-each-other/. [Accessed: Aug. 6, 2024].

[11] "Deep Blue Algorithm," *Professional AI*, [Online]. Available: https://professional-ai.com/deep-blue-algorithm.html. [Accessed: Aug. 6, 2024].

[12] "Deep Blue," *Stanford University*, [Online]. Available: https://stanford.edu/~cpiech/cs221/apps/deepBlue.html. [Accessed: Aug. 6, 2024].

[13] K. Wakefield, "A guide to the types of machine learning algorithms and their applications," *SAS Insights*, [Online]. Available: https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html. [Accessed: Aug. 6, 2024].

[14] "Unsupervised Learning: Algorithms and Examples," *AltexSoft*, Apr. 13, 2021. [Online]. Available: https://www.altexsoft.com/blog/unsupervised-machine-learning/. [Accessed: Aug. 6, 2024].

[15] "9 Real-Life Examples of Reinforcement Learning," *Santa Clara University Online Degrees*, Sep. 27, [Online]. Available: https://onlinedegrees.scu.edu/media/blog/9-examples-of-reinforcement-learning. [Accessed: Aug. 6, 2024].

[16] S. Grinschgl and A. C. Neubauer, "Supporting Cognition With Modern Technology: Distributed Cognition Today and in an AI-Enhanced Future," *Frontiers in Artificial Intelligence*, vol. 5, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.908261/full. [Accessed: Aug. 16, 2024].

11.0 Appendix

11.1 Revision Summary

Peer reviewers' insightful comments has led to a number of significant changes that have improved the report's depth and quality. A more thorough discussion of the main challenges, case studies, technological solutions resulted from feedback, as the draft was lacking in-depth detail with examples and citations.

Furthermore, an brief examination of the report's future implications for AI has been added in Section 5.0. This seeks to provide a viewpoint on the positive potential of AI use to enhance human reflection and decision-making, a key challenged expressed by the expert. Revision of problematic paragraphs, such the one about the impact of AI on personal freedom, addressed the earlier issue of run-on sentences. The report is now easier to read and understand, and complicated concepts are presented in a more concise manner thanks to this improvement.

A more thorough explanation of the different technology methods for addressing ethical issues was added to the section on technological solutions. Now, this section provides a more in-depth analysis of the real implementation of these ideas. More citations were added to the report to support its claims and increase the report's credibility. To further demonstrate the usefulness of the solutions mentioned, case studies and real-world examples were included, which increased the report's relevance and interest.

The lack of a table of contents was brought up in the comments, and it has now been added to enhance accessibility and ease of use in the report. Additionally, to better effectively display facts and key concepts, visual components have been integrated, such as Figure 1 in Section 3.1 and Table 1 in Section 3.4. These illustrations provide deeper understanding and support the content in a way that is closely related to the topics of their respective section.

The report has been enhanced by the changes made in response to the criticism from the peer reviewers. The report's clarity, relevancy, and overall quality have all been improved, and gaps have been addressed. The revision has been crucial in helping to improve the report. Not only has the report been improved in poor areas from peers, but the report as a whole has been modified to better address the grand challenges by Zittrain and have better structure.