Societal Benefits and Risks of Artificial Intelligence: A Succinct Survey

Sabrina Bergsten and Pablo Rivas
Department of Computer Science, School of Computer Science and Mathematics
Marist College, 3399 North Road Poughkeepsie, New York 12601, United States

Abstract— *In recent years we have learned that with great AI power comes great social responsibility. Thus, we studied the major societal benefits of AI in the areas of health care, education, and agriculture, discussing the positive change that AI has brought to each. We further address the risks society faces when AI is practiced without proper care, lack of responsibility, poor assessment of bias, and a complete disregard for ethically aligned designs. We aim to expose where we fall short so that we can move forward as a society.*

Keywords: artificial intelligence; responsible AI; ethics; AI for good; ethically aligned design

1. Introduction

Throughout history we have reserved the word intelligent for uniquely human skills and assets, "such as recognizing a familiar face, negotiating rush-hour traffic, or mastering a musical instrument" [1]. Within the last half century, however, this fact has quickly changed. Our most recent technological revolution has introduced a vast range of emerging innovations, one of which has been labeled Artificial Intelligence (AI). AI is a field of study within the computer sciences centered around the development of dynamic systems that can be taught to make decisions, predictions, and eventually independently learn new skills, within a certain context [2]. Applications utilizing AI have been used in a diverse variety of fields including health care, education, agriculture and business. The benefits AI brings to every area of study are countless and often lead to groundbreaking advancements. While it's common to only recognize the way this kind of technology pushes us forward in sectors like big data analysis or financial trend predictions, the amount of social good AI applications can bring about is just as remarkable. As Fig. 1 shows, many industries have been adopting AI technology in the recent years causing interesting changes in societies around the world.

The momentous innovation AI introduces has been helping lift societies to new heights and bringing essential growth to developing countries as well. As the orientation of AI into society as well as our everyday lives becomes more and more common, governments across the globe are working tirelessly to put laws and regulations in place that will uphold the integrity of this technology and eliminate the level of distrust this new intelligence so often brings with it.

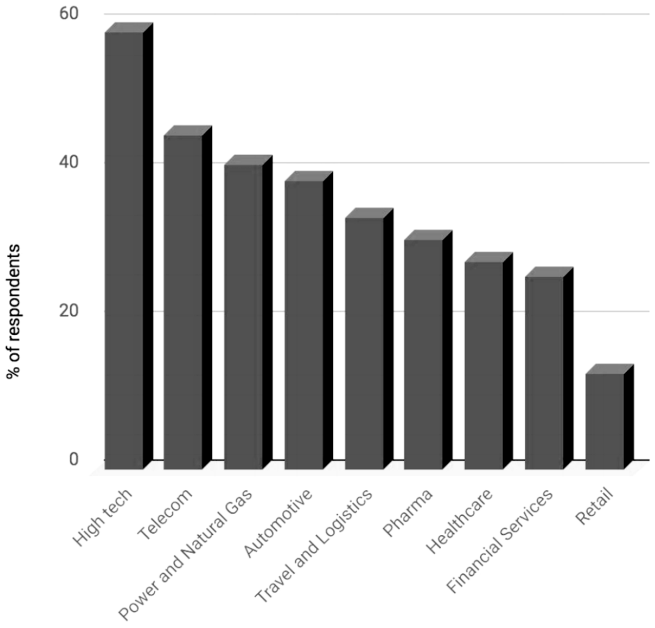

Fig. 1: Representation of the adoption rate for AI in product development by industry. Data retrieved from [3].

This short survey paper describes the areas in which AI can have large impact in Section 2; while in Section 3 we describe the potential risks of AI for societal systems. And finally, conclusions are drawn in Section 4.

2. Areas

We want to acknowledge at this point that while many of the following cases that we will discuss involve specific machine learning solutions to sever societal problems in different parts of the world, we will remain using the term AI to embrace all types of machine intelligence. The following paragraphs explain how AI has played pivotal, positive, role in the world in which we live. We begin discussing an important area: health care.

2.1 Health Care

Cripplingly low life expectancies as well as insufficient health care resources are omnipresent concerns in the developing world, where many treatable conditions go undiagnosed and, "infectious disease outbreaks frequently overwhelm available infrastructure" [2]. The introduction of AI into the Global South (GS) and across the world has been providing aid throughout these hardships by presenting an ever growing collection of new available resources. It's, "filling gaps in human expertise, improving the productivity of available health care workers, and enhancing disease surveillance" in societies that so desperately need it [4]. An example of this is displayed through the Dengue Fever (DF) epidemic of 2011 in Punjab, Pakistan. DF, a mosquito-borne viral infection, affects over 100 million people worldwide and over the past decade has consistently been spreading throughout Pakistan. In 2006, the first DF epidemic was reported in Punjab with a confirmed 113 cases and since then, a decade-wide analysis of the limited published data shows a threatening pattern of growing DF cases throughout Pakistan [5]. By 2011, there were a reported 16,580 confirmed cases of dengue viral infection and 257 deaths in the city of Lahore alone as well as nearly 5,000 cases and 60 deaths reported from the rest of the country [6]. A large part of this growing epidemic is the lack of documentation and data surrounding it. In 2011, after Pakistan was hit with this devastating outbreak, "a disease surveillance system was developed in Punjab to provide early warning of future outbreaks as a response." The resulting Punjab Intelligent Disease Surveillance System uses learning algorithms based on statistics to analyze data from a dengue hot-line and online news sources. This AI system offered local hospitals and government agencies real-time outbreak tracking with an astonishing level of geographic detail [7]. AI was able to provide a number of lifesaving health care resources without the added cost of hospital renovations to fit new treatment technologies or the cost of training local health care professionals, assuming these jobs would not have been outsourced to doctors from the United States or other western countries.

AI in health care is not limited to the progress it's making throughout the GS, but to nearly every facet in the area of study as well. AI driven research surrounding different disease types have lead to critical breakthroughs in search for a cure as well as a better understanding of the way they affect the body. The three main disease types AI literature encompass are cancer, nervous system disease, and cardiovascular disease [8]. AI cancer research includes the analysis of clinical images to identify skin cancer sub types as well as IBM Watson being utilized for oncology to assist with the diagnosis of cancer [8]. AI systems in neurology have been used to restore the control of movement in patients with quadriplegia [9]. It has been discussed that this technology's application in cardiology has the potential to diagnose heart disease simply through cardiac images [10].

It is object of study how much AI will change our way of life when it comes into contact with health care [11]; however, the education sector is also one that also has a great potential of positively affecting our way of life, as discussed next.

2.2 Education

The implementation of AI in education has completely remodeled the way the world views learning as well as opened doors to education, for so many, that are typically closed. The introduction and development of AI in various education systems throughout the world has only taken place throughout the last 30 years. Originally a large majority of education based AI technology was created for students with disabilities. Examples of these assistive technologies include, text to speech, speech to text, zoom capacity, predictive text, spell checkers, and search engines [12], [13]. These technologies now seem common, as many people utilize them in their everyday lives through personal computers. The frequency at which we use such tools can make us quick to forget the capacity at which they increase learning interactions of students around the globe, augment teaching possibilities, and enhance the designs of so many educational experiences [13]. AI in education has the power to grant access to practically limitless information and educational resources. With the use of AI, students of all levels of ability are being given the opportunity of personalized learning, while teachers are given access to resources relevant to each individual learner.

Intelligent Tutoring Systems (ITSs) are a product of AI technology that are built to adapt and personalize a learning experience to its user [2]. These systems, also known as cognitive tutors, are recreating one-on-one instruction without the cost or access of an on-the-ground personal tutor or teacher [2]. The spectrum of learning styles is vast and variable, and each individual fits on it in their own unique way. ITSs are using AI to find students' places on this spectrum and using these individual learning patterns to teach lessons with efficiency and effectiveness. One form of ITS is known as effective tutoring systems. These specific systems integrate emotional recognition to effectively enhance the tutoring adaptation to each specific student [14]. ITSs can even apply natural language processing to help identify mistakes in language and adapt to different levels of speech ability.

While the potential behind ITSs is clearly immense, there is already documented data supporting the fact that these systems have been successful in their pursuit to improve student performance. In [15], Kulik, *et.al.*, said that:

> *"A 2016 review of studies on the effectiveness of ITS found substantive improvements in performance in 78% of the 50 studies reviewed."*

There has also been strong evidence documented to suggest ITSs' potential success in the GS despite existing barriers [16]. If implemented correctly, ITSs will not only be an effective way to improve education, but an inclusive one as well.

While education is essential, agriculture is restricted to the parts of the earth that facilitate it. In the next paragraphs we explore how AI has also impacted this field.

2.3 Agriculture

The effect AI has begun to play in agriculture across the globe is not solely impacting farming industries, but whole economies and civilizations of people as well. In many countries in the GS specifically, much of a country's population is heavily reliant on farming as a source of food. The risk this poses comes with the many diseases, insect infestations, and droughts that threaten healthy crops each year [2]. Recent AI systems have implemented new methods of monitoring and detecting these risks before they occur as well as providing solutions to the teams of agricultural specialists and labourers that these populations of people so greatly rely on. The research and expert knowledge behind identifying crop disease is a niche field of study and experts are often unavailable locally, or financially, to large agricultural plantations in the GS. This is why researchers specialized in AI across the world have begun creating applications that farmers facing such threats in remote locations can use with just a click of their phone camera. AI researchers from Makerere University, for example, have optimized a similar method of using sample images taken on cell phones to detect crop disease. First, images of possible disease symptoms in the crops are collected and eventually fed into a machine learning algorithm, "to provide diagnosis and feedback, and a mapping function can be used to assess and prevent disease spread over time" [4].

3. Societal Skepticism and Risks

The implementation of AI in health care, education, and agriculture has proven to be extensively beneficial; however, there is still a major sense of skepticism and distrust in technology throughout society. Given the power that AI has to ignite change in almost every aspect of our lives, it makes sense that there is skepticism surrounding its implementation. This great power for change does not come without responsibility.

3.1 Responsible AI

The responsibility directly falls in the hands of computer scientists, machine learning scientists and practitioners, data scientists and engineers, and anyone involved in the development of AI technology. Their responsibility is to ethically and responsibly design and deploy AI technology. Unfortunately, to some extent, such responsibility also lies in the hands of our governments, to make sure that ethical practices have been followed when deploying AI systems for governance. The most renowned AI groups in Canada have paved the way showing that they are committed to practice responsible AI [17], leaving governmental institutions to take the next corresponding step.

In an ideal world, where everyone exercises responsible and ethical behaviour on the things we do and say for the sake of living in community, AI technology deployment could come without a second thought, and our society would be pushed to be that much greater. Alas, our world is not perfect. There are many risks involved in utilizing AI to its fullest potential. Some of these risks include learned bias, targeted misinformation, uncontrolled surveillance leading to a privacy concerns, and full automation leading to job losses without replacement. We shall now briefly address the risk posed by two specific and dangerous types of biases: Learning and Design [2].

3.2 Design and Learning Biases

As many popular studies show, humans are naturally born and raised surrounded by bias [18]. This leads us to manifesting our own implicit biases. Implicit biases can be hard to detect, yet they exist in all of us. Regardless of how consciously we attempt to be non-partisan, the biases we carry will always leave an effect on the work we produce, and data can prove this [19]. This is where such a large difference in humans and the machines we create lie. "Computers and algorithms are symbols of efficiency and accuracy. As machines and digital code, they are expected to function in a fair and unbiased way" [2]. The err in this is that while computers themselves strictly and without bias follow instructions, it is humans, filled with implicit biases, that are the ones programming these instructions into them. So often, the data we collect and analyze is littered with bias-creating factors, whether that is gender, ethnicity, financial background, etc. This kind of man made bias in technology has an official name: *biased computer systems*. Biased computer systems were originally defined by Friedman and Nissenbaum in 1996 as systems that "systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others" [20].

Further, when it comes to biased computer systems we have to differentiate between two types of biases that exist: *Design Bias* and *Learning Bias*. A *design bias* refers to programs, apps, tools, etc., that are initially designed with the implicit bias carried by the developers, whether they are aware of it or not. An example of this might include, an all-Caucasian male engineering team creating a voice recognition app. Since this team is so lacking in diversity, the app will recognize caucasian male voices successfully; however, the design bias percolated into the software leads to its low prediction rates for other voices, say, for example, voices of females. Examples like this have been consistently observed over and over again in varying fields. A field study

by James Rodger and Parag C. Pendharkar, demonstrated that medical voice-dictation software performs significantly better for men than women, and a design bias is most likely the cause [21].

A *learning bias* on the other hand, is a bias that occurs when a machine learning algorithm is trained on biased data sets and thereafter learns the patterns of the bias inherent in the data [22]. The evidence that AI applications are unfairly affecting important aspects of people's lives through biased data and programming already exists. This type of bias has can be prevented by exercising care when studying the data and performing the proper evaluations of performance for different groups [23]. This can be one the simplest biases to fix and prevent; however, it has the potential to be the most harmful for under-sampled demographics and protected groups. These types of bias are the leading cause of skepticism and prevent AI technology from advancing to its full potential in every-day-life scenarios, and thus, it is imperative that scientist, practitioners, and governments, exercise responsible AI to avoid hindering the increasing progress it has at the moment.

4. Conclusion

The greatest gift AI can give is access. Regardless of gender, ethnicity, ability, or background, every person is worthy of the same access to health care, education, clean water, and nutrition. AI uses the power of machine learning algorithms to give people across the globe an unprecedented opportunity to live better, healthier, and smarter. However, the deployment of AI in societies where resources are scarce continues to be a problem, especially in areas like health care. As Fig. 1 demonstrates, the rate of adoption of AI in the High Tech and Telecommunication industries is much higher than that of the Health Care and Pharmaceutical industries [3]. This could be for many reasons, two of which are well-known: lack of government support and lack of funding. While the lack of funding is motivated by corporate business models and capitalism, the support from government has to do with the problem of people's skepticism and lack of trust. Our best approach to increase trust and lower skepticism is to educate. First, educate people on what AI can and cannot do, ans then, educate our scientists to exercise responsible AI, to learn and apply fundamental principles of ethics, and to implement ethically aligned designs.

Acknowledgements

This research was supported in part by Marist College's Honors Program and the School of Computer Science and Mathematics.

References

[1] S. J. Russell and P. Norvig, *AI: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[2] M. L. Smith, S. NewPane, and C. Mendonca, "Crafting papers on machine learning," in *Proc. of the 17th International Conf. on Machine Learning (ICML 2000)*. Morgan Kaufmann, 2018, pp. 5–122.

[3] M. Chui and S. Malhotra, "Ai adoption advances, but foundational barriers remain," *Mckinsey and Company*, 2018.

[4] J. Quinn, V. Frias-Martinez, and L. Subramanian, "Computational sustainability and ai in the developing world," *AI Magazine*, vol. 35, no. 3, pp. 36–48, 2014.

[5] Z. Haider, F. Z. Ahmad, A. Mahmood, T. Waseem, I. Shafiq, T. Raza, J. Qazi, N. Siddique, and M. A. Humayun, "Dengue fever in pakistan: a paradigm shift; changing epidemiology and clinical patterns," *Perspectives in Public Health*, vol. 135, no. 6, pp. 294–298, 2015.

[6] E. A. Khan, "Dengue fever: A major public health problem of our time," *Rawal Medical Journal*, vol. 41, no. 2, pp. 139–141, 2016.

[7] T. Ahmad, N. A. Rehman, F. Pervaiz, S. Kalyanaraman, M. B. Safeer, S. Chakraborty, U. Saif, and L. Subramanian, "Characterizing dengue spread and severity using internet media sources," in *ACM DEV*, 2013.

[8] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Ai in healthcare: past, present and future," *Stroke and vascular neurology*, vol. 2, no. 4, pp. 230–243, 2017.

[9] C. E. Bouton, A. Shaikhouni, N. V. Annetta, M. A. Bockbrader, D. A. Friedenberg, D. M. Nielson, G. Sharma, P. B. Sederberg, B. C. Glenn, W. J. Mysiw, *et al.*, "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, no. 7602, p. 247, 2016.

[10] S. E. Dilsizian and E. L. Siegel, "Ai in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment," *Current cardiology reports*, vol. 16, no. 1, p. 441, 2014.

[11] J. H. Chen and S. M. Asch, "Machine learning and prediction in medicine-beyond the peak of inflated expectations," *The New England journal of medicine*, vol. 376, no. 26, p. 2507, 2017.

[12] S. Ö. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, *et al.*, "Deep voice: Real-time neural text-to-speech," in *Proc. of the 34th International Conf. on Machine Learning*. JMLR. org, 2017, pp. 195–204.

[13] S. A. Popenici and S. Kerr, "Exploring the impact of ai on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, p. 22, 2017.

[14] S. Petrovica, A. Anohina-Naumeca, and H. K. Ekenel, "Emotion recognition in affective tutoring systems: Collection of ground-truth data," *Procedia Computer Science*, vol. 104, pp. 437–444, 2017.

[15] J. A. Kulik and J. Fletcher, "Effectiveness of intelligent tutoring systems: a meta-analytic review," *Review of Educational Research*, vol. 86, no. 1, pp. 42–78, 2016.

[16] B. D. Nye, "Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context," *International Journal of Artificial Intelligence in Education*, vol. 25, no. 2, pp. 177–203, 2015.

[17] "Montreal declaration for a responsible development of ai," *Announced at the conclusion of the Forum on the Socially Responsible Development of AI*, 2017. [Online]. Available: https://www.montrealdeclaration-responsibleai.com/the-declaration

[18] I. Misra, C. Lawrence Zitnick, M. Mitchell, and R. Girshick, "Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2930–2939.

[19] J. Holroyd, R. Scaife, and T. Stafford, "Responsibility for implicit bias," *Philosophy Compass*, vol. 12, no. 3, p. e12410, 2017.

[20] B. Friedman, *Human values and the design of computer technology*. Cambridge University Press, 1997, no. 72.

[21] J. A. Rodger and P. C. Pendharkar, "A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application," *International Journal of Human-Computer Studies*, vol. 60, no. 5-6, pp. 529–544, 2004.

[22] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356, no. 6334, pp. 183–186, 2017.

[23] P. Rivas, "Ethical machine learning: Identifying and removing unfariness caused by biased data," *New York Celebration of Women in Computing Conference*, 2019.