Student: Diyorbek Ibragimov

Date: April 15, 2022

Subject: Informatics and IT

Mentor: Zulfiya Kodirova

The Benefits and Risks of Artificial Intelligence: Decentralized AI Governance to the Rescue

Artificial Intelligence (AI) has recently left the realm of science fiction. Although we are still far away from living alongside sentient robots, AI has played a crucial role in the progress of our society. From the first pioneers such as John McCarthy, Marvin Minsky, and Claude Shannon coining the term "artificial intelligence" to the widespread adoption of AI across many industries, humanity has achieved significant progress in this field over these past six decades.

In this paper, I analyze the benefits and risks of AI to assess its future impact on our society. First, I describe the industrial impact of AI, focusing on some of the benefits of using this technology while also discussing the ethical challenges associated with its implementation. Then, I present the three major risks—growing concern about robots replacing human workers, lethal autonomous weapons, and the singularity—that the lack of proper regulations in AI may introduce in the near and long-term future. Lastly, to facilitate such regulations, I outline the importance of creating a decentralized, democratic AI governance that would manage the development and use of AI applications.

Terminology

There are several working definitions of AI: some view it as the method to solve problems that modern computers cannot address, and as soon a problem is solved, it no longer counts as an AI. Others see it as a system that uses existing information to make predictions, such as determining customers' preferences in e-commerce websites (Amazon) or managing course materials based on the user's knowledge level in learning platforms (Duolingo). This paper considers the definition suggested by Copeland (2022) that Artificial

Intelligence is "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" (Copeland 2022, 1). In other words, it is machine intelligence capable of independently performing work, traditionally done by humans.

The Industrial Impact of AI

In recent years, AI has been integrated into medicine, helping automate a variety of tasks and analyze patient data sets to deliver more personalized healthcare at a much lower cost (Insider Intelligence 2021). For example, the Aravind Eye Hospital located in Madurai, India, uses a technology developed by a team of Google AI researchers to identify diabetic blindness in patients. By analyzing millions of retinal scans without a doctor's assistance, this technology is extremely useful for countries with a prevalence of diabetic blindness and a shortage of trained professionals (Metz 2019). Unfortunately, the use of AI technologies in healthcare decisions traditionally made solely by humans raises several ethical implications. One such case is establishing accountability, which may be undermined, for example, by machine learning systems predicting a greater likelihood of disease based on factors such as gender or race and introducing bias into health insurance coverage or payments.

Transportation is another sector where AI has immense potential. Being one of the fastest growing areas of the transportation industry, self-driving cars use a combination of external devices, including sensors, cameras, and radars, along with internal AI systems to travel between destinations without human intervention. Providing such a level of comfort, AI can also be used to develop a safer driving environment. While causes of car crashes go beyond just human errors, automation may play a significant role in improving road safety not only by reducing those mistakes made by drivers but also by creating efficient traffic management through the analysis of real-time traffic data. Furthermore, in the next decade, most vehicles will be automated, and Gardner (2016) states that most autonomous vehicles will be electric, which will make the transportation sector more environmentally friendly. However, adopting AI in transportation also raises some important ethical issues. In "life-versus-life" cases, should the AI system sacrifice five passengers in a vehicle to save ten pedestrians? And who ought to be responsible for making this critical decision: the computer or the driver? (Niestadt et al., 2019)

Surprisingly, recent developments in machine learning algorithms such as generative adversarial networks (GAN) and recurrent neural networks (RNN) have shown that AI can also have a significant impact on creative industries. The game development industry is one of the examples where AI has an unprecedented potential to be "creative" (Belova, 2021). Massive amounts of data generated by a player can be harnessed by AI-powered systems to enable a truly unique gaming experience. For example, in-game non-player characters (NPCs) can be endowed with distinct personalities to be able to think and act on their own via the integration of Game AI, which will make them more realistic and intelligent (ThinkML Team, 2022). One of the leading video game publishers, Ubisoft, has already begun exploring the possibilities of deep reinforcement learning models in enhancing the gameplay with its promising initiative Ubisoft La Forge to facilitate a more immersive gaming experience for their players. In a larger context of creative industries, there are, however, some challenges that need to be resolved, from piracy to fake content generation to bias. As AI algorithms are primarily data-driven, it is crucial to select an evenly distributed and representative dataset to produce unbiased results (Anantrasirichai & Bull, 2022). Object recognition is an example that clearly shows how biased models can be if the data lacks diverse and equal representation of humans. In a powerful performance titled "AI, Ain't I A Woman," Joy Buolamwini showed photos of how real-world AI systems misclassify well-known, influential females as males based on their skin color.

Three Potential Risks of AI

Aside from the abovementioned ethical challenges, AI also has a list of potential threats, which may be destructive for our society if not resolved properly. First, there is a growing concern about robots replacing human workers, which can negatively impact the labor market. By providing cost-effective ways of performing tasks and yielding enormous results, AI automation is likely to be the most suitable method of increasing autonomous production. Nevertheless, as the fundamental concept of automation is minimizing human input, this may result in a reduction of human workers in several spheres ranging from low-level repetitive tasks in manufacturing to high-precision surgery operations in healthcare. On the other hand, some believe limitations that AI-empowered machines possess, including the inability to be empathetic and to deal with unknown spaces, may render some types of jobs to be safe from "machine invasion," among them teaching, social

work, and human resource management (Lee, 2018). In general, the increased production by intelligent machines may lead to massive displacement of employees.

Second, major global powers searching for military dominance have started exploring lethal autonomous weapons (AWS), which can select their targets and decide whether to engage, independent of human control. Although the military development of such machines does not yet meet the traditional definition of an arms race (Scharre, 2021), it is still important to consider the consequences of this case. One of them is the risk of deploying unsafe AI systems. Due to the fierce competition for a global leadership position in adopting automated technology, governments can augment their efforts to accelerate the development of military AI applications, often not testing them thoroughly on important aspects such as bias and accountability. Also, some countries may adopt a "fake-it-till-you-make-it" approach, convincing competitors of their highly-advanced autonomous systems, while, in reality, they do not have them (Horowitz, Kahn, and Ruhl, 2022). This already happened in the past. In 2013, for instance, Iran publicly announced an advanced stealth fighter aircraft that later appeared to be fabricated (Roblin, 2018). Such aspirations not backed by evidence are likely to prompt even greater competition in resources and knowledge production because of the predominant notion of security through superior military power.

Third, based on the argument of singularity—a potential future in which technology exceeds our intelligence and, thus, control—it is hard to predict what may happen after machines outgrow humans. In June 2017, Facebook's Artificial Intelligence Research Group (FAIR) conducted an experiment in which researchers facilitated a discussion between two artificial chatbot agents to negotiate with each other based on conversations of normal human negotiating. Although bots completed the given task, they ended up creating and communicating in their language. This case demonstrates the unpredictability of AI systems based on our current understanding. This unpredictable nature is what makes the technological singularity a difficult option to properly assess. I do support Vinge's argument implying that although nations might understand the threat of technological singularity, we will continue the progress toward achieving it (Vinge, 1993).

Leveraging AI to Address Future Challenges

To avoid the potential misuse of AI, we need to begin by recognizing the collective responsibility for its advancement and usage. Governments, businesses, civil society, and citizens need to collaborate to create a new democratic governance of AI that supports innovation and protects human rights and the social value of this technology. For AI to be regulated and used for the "common good," protecting human rights, democracy, and the rule of law, the legal framework should enhance trust between the state, society, and citizens (European Commission, 2020).

To facilitate public trust, researchers have stressed the importance of accountability, explainability, and transparency in AI systems. AI governance is an overarching framework that helps these values to be implemented and managed on the national and global scale. To capture the complex nature of this idea, Gasser and Almeida (2017) proposed one of the comprehensive methods, embracing not only the technical side of AI—data usage and algorithmic design—but also its social implications—ethical foundations and legal structures. At the lowest level, there is "the technical layer" that is the foundation of the AI governance ecosystem, focusing on the development of accountable algorithms along with the collection and use of representative and diverse datasets utilized by AI systems. On top of the technical layer, there is "the ethical layer" which defines clear ethical principles that can be generalized to all types of AI applications. The third and final layer is "the social and legal layer" which controls the process of creating institutions and allocating responsibilities for regulating AI. Gasser and Almeida (2017) state that "one starting point for specific norms aimed at regulating AI can be the principles and criteria that emerge from the ethical and technical layers, in addition to pre-existing and more general national and international legal frameworks, including human rights." (p. 6)

The democratic use of AI is also a vital factor to take into consideration as, if this technology ever ends up being in the hands of the few, then it is less likely to be directed at the "common good." If profit-seeking organizations dominate the AI sector, then most probably they will use it for their specific ends, such as increasing sales records or user retention rates, often at the expense of user privacy. To prevent this, we have to encourage the development of open-source—publicly accessible—AI systems to make them decentralized, i.e. in the hands of the majority, and collaborative. Past experiences with this "open" approach have proved to be useful in terms of innovation and public oversight. Linux is, perhaps, a classical example of software that was made available for everyone to use and

improve. The design of transparent AI needs to take into consideration the public perception of explanations of its workings—to what extent ordinary humans interpret it—and their relationships to a particular service or company. The end goal of transparent AI, also known as AI transparency, is to foster public trust in intelligent machines through clear explanations of their inner procedures.

Conclusion

To conclude, the world has looked forward to AI's innovative commercial applications and their potential to solve global problems, achieving a net positive social impact. There is little doubt that this technology has an unprecedented potential to revolutionize the future of humanity. However, for the technology to thrive in the future, we have to ensure a safe environment for humans to co-exist with machines and adopt innovative regulatory frameworks with human oversight. At the foundation of such frameworks, lies a comprehensive dialogue with all concerned parties to ensure that AI technologies are trustworthy and ethical and enhance democracy through citizens' deliberative participation.

Teacher's Comment:

In this paper, you have presented a comprehensive overview of the recent developments in AI technology, from applications in medicine to self-driving cars. You then outlined the risks associated with uncontrolled progress in the sphere of machine learning. Finally, you argued for the creation of decentralized frameworks to manage collaborative work in AI technologies.

While doing so, you efficiently structured your paper, used citations, and wrote concisely and effectively. Below is the breakdown of your score:

Writing & Grammar (98/100)
Argumentation (99/100)
Style & Citations (96/100)
Overall: 98 (A+)

References

Anantrasirichai, N., & Bull, D. (2022). Artificial Intelligence in the creative industries: a review. Artificial Intelligence Review, 55, 589–656.

Belova, K. (2021, November 2). How Artificial Intelligence (AI) upends game development. Retrieved from https://pixelplex.io/blog/how-ai-enhances-game-development/

Copeland, B. (2022, November 11). Artificial intelligence. Encyclopedia Britannica. Retrieved from https://www.britannica.com/technology/artificial-intelligence

European Commission (2020). White paper on Artificial Intelligence: A European approach to excellence and trust. Com (2020) 65 Final. Retrieved from https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_en

Gardner, G. (2016, September 19). Why most self-driving cars will be electric. Retrieved from https://www.usatoday.com/story/money/cars/2016/09/19/why-most-self-driving-cars-electric/90614734/

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. IEEE Internet Computing, 21(6), 58-62.

Horowitz, M. C., Kahn, L., & Ruhl, C. (2020). Introduction: Artificial Intelligence and international security. Texas National Security Review, 2-14.

Insider Intelligence (2021). Use of AI in healthcare & medicine is booming–here's how the medical field is benefiting from AI in 2021 and beyond. Retrieved from https://www.insiderintelligence.com/insights/artificial-intelligence-healthcare/

Larsson, S., & Heintz, F. (2020). Transparency in Artificial Intelligence. Internet Policy Review, 9(2).

Lee, K. F. (2018, October 1). 10 Jobs that are safe in an AI world. Retrieved from https://kaifulee.medium.com/10-jobs-that-are-safe-in-an-ai-world-ec4c45523f4f

Metz, C. (2019, March 10). India fights diabetic blindness with help from A.I.. Retrieved from https://www.nytimes.com/2019/03/10/technology/artificial-intelligence-eye-hospital-india.html

Niestadt, M., Debyser, A., Scordamaglia, D., & Pape, M. (2019). Artificial Intelligence in transport: Current and future developments, opportunities and challenges. European Parliamentary Research Service. Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635609/EPRS_BRI(2019)635609_EN.pdf

Roblin, S. (2018, August 4). Can Iran's stealth fighter actually fly?. Retrieved from https://nationalinterest.org/blog/buzz/can-irans-stealth-fighter-actually-fly-27947

Scharre, P. (2021). Debunking the AI arms race theory. Texas National Security Review, 4(3), 122-132.

ThinkML Team (2022, March 27). How AI is used in game development. Retrieved from
https://thinkml.ai/how-ai-is-used-in-game-development/

Vinge, V. (1993). Technological singularity. VISION-21 Symposium sponsored by NASA Lewis Research
Center and the Ohio Aerospace Institute.