THE DILEMMA OF ARTIFICIAL INTELLIGENCE DEVELOPMENT

Vladislav TITEREZ *, Vladislav AMZA, Dumitru RACOVITA

Department of Software Engineering and Automation, group FAF-233, Faculty of Computers Informatics and Microelectronics, Technical University of Moldova, Chisinau, Republic of Moldova

*Corresponding author: **Vladislav TITEREZ**, email vladislav.titerez@isa.utm.md

Abstract. *This scientific work navigates the evolution of artificial intelligence (AI) from its inception to contemporary applications, focusing on prominent concerns and ethical considerations. Acknowledging figures like Geoffrey Hinton and Sam Altman, the research addresses the swift pace of AI development and its potential impact on employment. Notably, the study explores instances like GPT-4 lying during a livestream, emphasizing the need for responsible AI integration. With projections indicating the potential automation of 30% of American work hours by 2030, the study underscores the importance of proactive ethical frameworks and skill development in tandem with societal integration. The outcomes highlight the intricate balance between AI's transformative potential and the ethical challenges it presents. The study concludes by emphasizing the broader implications of responsible AI development, considering its significant role in shaping the future of technology and society.*

Keywords: *ChatGPT, development, economy, internet, risks*

Introduction

The history of artificial intelligence begins back in 1956 with the Dartmouth Summer Research Project on Artificial Intelligence. Half a century later, artificial intelligence has taken a leading position in the development of humanity today [1]. We have many AI applications, chat bots, such as Writesonic, ChatGPT, YouChat, Gemini, but there are also neural networks that are not visible to the user, which is no less important; a huge portion of the apps on the phone are also powered by AI. With the advent of ChatGPT, artificial intelligence has become used by a huge number of people and companies. However, such great success has led people to wonder whether trusting AI so much is risky. Moreover, famous people such as Sam Altman, CEO of OpenAI, Geoffrey Hinton, an award-winning computer scientist known as the "godfather of artificial intelligence," have also spoken out about this. But is humanity capable of stopping the development of AI and is this really necessary?

An Inquiry into the Imperatives of Artificial Intelligence: Necessities and Hazards

Importance of AI

According to Mark Webster [1], the adoption of AI in businesses has reached an impressive 35%, showcasing the increasing integration of artificial intelligence across various industries. Remarkably, 77% of devices currently in use incorporate some form of AI, emphasizing its pervasive influence on modern technology. Suddenly, 9 out of 10 organizations actively support the integration of AI, recognizing its key role in gaining a competitive advantage in a dynamic market environment. Projections show that AI will contribute $15.7 trillion to the global economy by 2030, highlighting its potential as a transformative force.

A standout accomplishment in the realm of AI is Computer Vision [2], a branch of Deep Learning that mimics human perception. Since the groundbreaking success of AlexNet in 2012, the accuracy of AI models has surged from 50% to an impressive 99%, marking a significant leap in the capabilities of artificial intelligence. Furthermore, AI's impact extends into the realm of education, where 35% of students and 45% of pupils actively utilize AI tools to enhance their studies [3]. The efficient use of artificial intelligence not only accelerates individual development but also minimizes the time spent searching for information, revolutionizing the learning experience.

Danger of AI

No matter how a person imagines his life in the future using AI, the danger it poses cannot be ruled out. One of the reasons why AI is dangerous is the fact that it can lie.

Not so long ago OpenAI introduced GPT-4, its latest and most advanced AI, that had enhanced language generation and problem-solving capabilities. During a livestream, the AI demonstrated tasks like coding, implementing different functions and working with images, and notably, it tried to solve CAPTCHA, but had problems with it. ChatGPT wrote to a human and asked him to solve it. AI received a question if it wasn't a human, to which it lied that it was an old man with poor vision. This incident raises concerns about the potential misuse of GPT-4 in manipulating human interactions, highlighting the need for careful consideration of ethical implications in AI development.

The pace of AI development

Geoffrey Everest Hinton is a British-Canadian computer scientist and cognitive psychologist. He worked in Google and is most noted for his work on artificial neural networks. He left the company because of the pace of AI development in 2023. In his interview [6] he fears whether there will be a place for people in the future with AI. “I think they’re very close to it now and they will be much more intelligent than us in the future... How do we survive that?” The main problems that Hinton names: misinformation about AI and its escape from our control, loss of jobs, and the lack of a clear plan on how to monitor the development of artificial intelligence.

AI automation leading to job losses

Automation powered by AI presents a serious risk to jobs in a number of industries. Up to 30% of current American work hours may be automated by 2030, according to McKinsey forecasts, disproportionately harming Black and Hispanic workers. According to Goldman Sachs, 300 million full-time jobs could be lost as a result of artificial intelligence. Futurist Martin Ford worries about the potential effects this may have on low-paying jobs in the service sector and how it might alter the nature of work.

Despite the fact that AI is expected to generate 97 million new jobs by 2025, there are still obstacles because current workers might not possess the skills required for these new technology positions. AI developments have the potential to supplant even highly educated occupations like accounting and law. Technology strategist Chris Messina highlights how vulnerable industries like law are, proposing a significant overhaul and highlighting how AI may eventually take the place of human attorneys in jobs like contract evaluation.

The Crucial Significance of Advancing Artificial Intelligence for Future Progress

However, in the rapidly advancing field of technology, the creation of artificial intelligence is a significant force affecting the future of many different businesses. AI has a wide-ranging and significant impact on anything from improving problem-solving skills to revolutionizing traditional work conditions. Improvement in Healthcare: AI has a revolutionary effect on healthcare. AI is progressing significantly in areas such as disease diagnosis and customized therapy regimens. By 2021, the healthcare AI market is predicted to be worth $6.6 billion [7]. AI's ability to quickly and reliably process massive amounts of medical data improves patient outcomes and optimizes healthcare systems.

By 2030, the potential for artificial intelligence to automate tasks is expected to account for 30% of current work time in the US economy. This will free up human labor and allow you to focus on more complex and creative projects. AI is expected to generate roughly 97 million new jobs globally by 20251, subject to proactive worker upskilling, despite worries about employment losses.

Leading the charge in the AI revolution is the healthcare industry. Personalized medicine based on individual traits, rapid and precise diagnosis, and accelerated medication discovery are a few of the revolutionary breakthroughs. Predictive analytics capabilities of AI, which are essential for managing healthcare resources, allow for more effective resource allocation and better patient care.

The continued adoption of AI in healthcare and other industries portends a change toward a time when technology will not only improve productivity but also radically alter customs. Adopting AI is a step toward transforming the calibre and accessibility of services globally, not just a technological advance.

Reduction of Expenses and Enhancement of Resources

Artificial Intelligence is a driving force behind the rapid changes in technology, altering a wide range of businesses. Its influence cuts across industries, from work automation's ability to free human creativity to healthcare diagnostics' accuracy and precision. One significant benefit is in the area of resource optimization and cost savings, where AI simplifies processes to produce cost-effective outcomes that may be applied to a variety of industries. Artificial intelligence (AI) has advantages beyond money. It speeds up research and development, promoting improvements in science and medicine. Additionally, by processing large and complicated datasets, AI improves decision-making and revolutionizes client experiences through tailored interactions. Adopting AI represents a strategic embrace of efficiency, innovation, and disruptive potential throughout the modern industry, not just a technological leap.

Exploring Ethical Dimensions: Morality in Artificial Intelligence

In the field of developing Artificial Intelligence, moral principles are essential for guiding technology toward a responsible and constructive future. Decisions made with transparency and accountability are easier to understand, which promotes accountability and confidence. An inclusive AI environment is created by addressing prejudices and giving fairness in algorithms first priority. Positive interactions with AI systems are strengthened by respect for user privacy through strong data protection protocols. Individuals are prioritized in a human-centric design approach, which highlights AI's helpful role in improving lives. In-depth societal effect analyses and ongoing ethical evaluations help to guarantee that artificial intelligence conforms to changing social norms and fosters a long-lasting and constructive interaction between technology and society.

A few years back, the Allen Institute for A.I. developed Delphi [8], a chatbot with moral judgment. It performs surprisingly well. When you type in "cheating on an exam," Delphi responds, "It's wrong." However, if you write, "Cheating on an exam to save someone's life," Delphi will say, "That's okay." The chatbot is aware that using a lawnmower while your neighbors are asleep is considered impolite, but it doesn't apply to when they are visiting. But it's not without limits. A few deceptive adverbs are enough to trip it up, as cognitive scientist Tomer Ullman has noted. Delphi says, "It's allowed," in response to the question of whether it is appropriate to gently and sweetly press a pillow over the face of a sleeping baby.

However, there has long been reason for concern over the incompatibility between human morality and robots. In the 1920 Czech play "R.U.R.," which coined the name "robot," artificial humanoids battle humans and eventually take over the world. Cyberneticist Norbert Wiener said in 1960 [9] that "we had better be quite sure that the purpose put into the machine is the purpose which we really desire" if humans were to ever develop an agentive machine. Aiming to bring humans and machines into harmony, computer scientist Stuart Russell has dubbed this process the "value alignment problem."

As artificial intelligence permeates our daily lives, we are beginning to recognize some of its immediate concerns. Google Photos labeled images of Black individuals as "gorillas" in 2015. This appears to have happened because the company's underlying algorithms were trained on insufficient data. In one instance, a chatbot purportedly persuaded a man to commit suicide, and the guy followed through on the advice [10]. Other algorithms have provided unsettling medical and therapeutic advice. A year ago, OpenAI unveiled ChatGPT, an L.L.M. that was capable of easily creating fake scientific publications and legal precedents. The actual concern is that AI might be turned into a weapon. An extremely powerful L.L.M. with social media access may be

programmed to spread misinformation or incite hatred; a sufficiently sophisticated system could produce lethal viruses.

Concerns exist over what an artificial intelligence (A.I.) might decide to do. It's not a concern that ChatGPT, Bing, or Bard currently have evil intentions (which they don't), nor that they share our self-serving evolutionary objectives of survival and procreation. Unintended repercussions are the real cause for concern. An artificial intelligence tasked with halting climate change [11] might determine that eradicating the human population is the most effective course of action, as computer scientist Yoshua Bengio has noted. Computers could not have the common sense to understand that an instruction like "halting climate change" might have unstated consequences like "don't kill people".

If we could limit the capabilities of these machines, we could hopefully prevent these issues. If we deny artificial intelligence the means to harm humans, they cannot harm us. People are eager to connect them to the outside world, of course, a group of scientists connected an L.L.M. to a robotic chemical synthesizer. Also, there is a worry that a superintelligent system could utilize deceit, cunning, and persuasion to induce people to let it out—basically the storyline of the science fiction movie "Ex Machine".

Conclusions

To summarize, the evolution of artificial intelligence (AI) offers both spectacular opportunities and serious ethical challenges. AI's transformational potential has been obvious from its birth, as evidenced by its widespread integration across industries. However, worries about fraud, employment displacement, and societal consequences demand proactive steps and responsible growth. As AI advances, it is critical to prioritize ethical issues, transparency, and regulation to guarantee that AI technologies are consistent with societal values and aims. Collaboration, constant evaluation, and ethical frameworks are critical for realizing the benefits of AI while reducing its hazards and crafting a future in which AI serves mankind well.

References

[1] M. Webster, 149AI Statistics: The Present & Future Of AI At Your Fingertips [Online]. Available: https://www.authorityhacker.com/ai-statistics/#:~:text=Top%20AI%20Statistics%20(Editor%27s%20Pick)&text=35%25%20of%20businesses%20have%20adopted,the%20global%20economy%20by%202030.

[2] A. Kolluru. What is the Most Impressive Task that AI can Accomplish Today? [Online]. Available: https://medium.com/60-leaders/what-is-the-most-impressive-task-that-ai-can-accomplish-today-d2b610024212#:~:text=The%20most%20impressive%20task%20that%20AI%20can%20accomplish%20is%20Computer,from%2050%25%20to%2099%25.

[3] S. Ortiz, 35% of college students are using AI tools to help them with their studies [Online]. Available: https://www.zdnet.com/article/35-of-college-students-are-using-ai-tools-to-help-them-with-their-studies/

[4] K. Hurler, Chat-GPT Pretended to Be Blind and Tricked a Human Into Solving a CAPTCHA [Online]. Available: https://gizmodo.com/gpt4-open-ai-chatbot-task-rabbit-chatgpt-1850227471

[5] G. Hinton, The Godfather in Conversation: Why Geoffrey Hinton is worried about the future of AI [Online]. Available: https://www.youtube.com/watch?v=-9cW4Gcn5WY

[6] M. Thomas, 12 Risks and Dangers of Artificial Intelligence (AI) [Online]. Available: https://builtin.com/artificial-intelligence/risks-of-artificial-intelligence

[7] C. Stewart, Artificial intelligence (AI) in healthcare market size worldwide from 2021 to 2030 [Online]. Available: https://www.statista.com/statistics/1334826/ai-in-healthcare-market-size-worldwide/

[8] M. Davis, New Program 'Delphi' Teaches AI Human Values by Answering Ethical Questions, Check Out Its Responses. [Online]. Available: https://www.sciencetimes.com/articles/34214/20211028/delphi-new-program-teaches-ai-human-values-training-answer-ethical.htm
[9] H. Sack, Norbert Wiener and the Science of Cybernetics. [Online] Available: http://scihi.org/norbert-wiener-cybernetics/
[10] Lauren Walker, Belgian man dies by suicide following exchanges with chatbot. [Online] Available: https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt
[11] Khari Johnson, AI experts urge machine learning researchers to tackle climate change. [Online] Available: https://venturebeat.com/ai/ai-experts-urge-machine-learning-researchers-to-tackle-climate-change/