ASA Research Note

EMERGING RISKS IN AI DEVELOPMENT

AUTHOR: Ananya Mishra

PUBLISHED: January 2024

WRITTEN: December 2022

KEYWORDS: Artificial Intelligence, new technologies; emerging technologies;

ABSTRACT: This paper discusses the risks of the creation, deployment, adoption, and expansion of the use of artificial intelligence, specifically during the development phase of its lifecycle. After examining some specific risks, the author explores potential risk mitigation policies that could aid organizations and governing bodies in navigating this rapidly changing technology environment.

Artificial intelligence (AI), a term first coined in the mid-1950s, has grown to have definitions and implications beyond an average human's imagination. Google defines AI as "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."[1] Companies have increasingly begun to adopt AI into their business processes, which exposes them to a pool of opportunities; however, with such opportunity comes a high level of risk, given that there is so much more at stake that must be accounted for. This paper discusses the risks that come with the adoption of AI, following up with mitigation policies that can aid in maneuvering the difficulties that such technology developments project.

AI can enhance decision-making processes by automating monotonous and time-consuming tasks, which boosts a firm's efficiency and allows for deeper dives into data generated through AI modeling. This enables an environment wherein decisions are highly informed and backed up by data that support the claims being made. By integrating AI with industrial processes, companies can increase their competitive advantage and accelerate growth.[2] Given the recent boom in the use and adoption of AI, companies across all kinds of industries seem to be in a rat race to implement this advancement in their business processes. However, what firms often forget is that nothing good comes without consequences. The rush to match up with competitors and the desire to be the best stems from a lack of consequentialism, accountability, and integrity.

One of the most significant flaws in the development of AI models is the algorithmic bias in the systems that stems from the biases of the people who are involved in its development.[3] People create the machine learning algorithms, and human bias is almost inevitably introduced as a result, either "intentionally or inadvertently."[4] It is in human nature to form opinions that lean toward a specific school of thought, which sometimes can prove to be unfair when looked at through a broader perspective. And these biases "don't simply exist in a vacuum or in our minds — they affect the way we make decisions and act."[5] Machine learning models must learn from humans – and so the risk exists that existing biases are passed into the AI model, are amplified, and negative patterns are reinforced. An example of such an occurrence was Amazon's recruiting tool. Although not designed intentionally, it was discovered that the tool held a bias against women and downgraded their resumes. This is because since "most of these resumes were submitted by men, the system taught itself to favor male candidates."[6] This demonstrates how easily an idea, which is very evident in our

society, can reflect onto a technological tool, causing serious concern for companies and, more so, individuals, as it creates room for injustice and unfairness.

The datasets used to train AI models are often another source of algorithmic bias. The quality and size of training data being used play a huge role in deciding the gravity of bias the results can bring forth. This practice is intertwined with the concept of human bias. This is because the ball lies in the court of the algorithm developers when deciding what datasets to use. Often, the data set chosen can cause skewed results that can lead to inaccurate predictions. The results are generally presented as an accurate representation of the solution to the issue at hand, however, more often than not, the results are an overrepresentation, which emphasize a specific idea or perspective.[7]

There have been ethical issues surrounding the collection of data used to train AI models as well. Several companies that aim at being able to build their AI models have been found guilty of collecting and using data to "train" those AI models in unethical ways. This is due to the lack of accountability and ethics in a firm's business practices. Innovation is a significant step to take when growing and expanding a company. Still, there is a need for companies to act responsibly when making use of data, which may affect an individual's right to privacy and the value of content that's made accessible. Collins has rightly described the five stages of decline.[8] The stage of the "Undisciplined pursuit of more"[9] aptly depicts the lack of ethics in companies. When companies begin to lose track of what is right vs wrong and are ready to use what's beneficial to them without considering the consequences, there is a high chance that this leads to their downfall or, in Collin's words, "Capitulation to irrelevance or death."[10]

One of the most recent cases of unethical use of data is by IBM. There was a class action lawsuit filed against the company for using photographer Janecky's photo library containing 99 million photos to create a database that would help "train" the model to reduce bias.[11] This seems like an idea with good intentions, but not when it is performed without the individuals' consent. Another similar case is that of Clearview AI's. They were found guilty of using images "scraped from social media and other internet-based platforms such as Venmo"[12] to create a facial recognition database.

This gives rise to the endangerment of intellectual property (IP) and its protection. Scraping and duplicating art and information on the internet leads to a decrease in the value of creativity by professionals, in addition to the violation of an individual's right to privacy. Such cases indicate that with the boom in AI, digital technology, and the Internet, copyright infringement is being made easier and cheaper.[13] "The United States is the world leader in the field of AI development, with new applications and capabilities being developed almost every day."[14] However, where the United States fails to lead is in the aspect of policy creation and law enforcement toward consumer protection and privacy. Although "Reasonable Security or information/data protection has been defined and redefined over the years"[15], current privacy laws still barely consider the growth in technology and AI's capability to generate new data breaches and privacy violations. This can be realized as a system risk[16], which can prove to be harmful not only to an individual but the nation as a whole.

Given the speed at which AI is growing, companies adopting this technology have a high chance of being exposed to cyber and third-party risks.[17] This is a result of a lack of in-house skills, which leads companies to outsource AI project management to third-party entities, which increases the probability of cyberattacks. Third-party involvement plays a huge role in creating a point of vulnerability in a company's internal system, which becomes an easy gateway for hackers.

While having multiple stakeholders, such as third-party vendors, in the deployment of AI in companies can be risky, having little to no human intervention in AI deployment can also cause huge risks. There has been

an overestimation of the capabilities of AI by humans, which leads to a lack of human oversight in a company's system and processes. This often results in programmatic errors and system failures.[18] AI systems only can perform tasks based on the training data and information inputted in the formation stage. Trusting the system to carry out activities without any human interference and monitoring can lead to unreliable results and serious consequences. One such case was the launch of Microsoft's chatbot - Tay. Microsoft had designed this chatbot as "an experiment at the intersection of machine learning, natural language processing, and social networks."[19] "Tay was designed to learn more about language over time... Eventually, her programmers hoped Tay would sound just like the Internet."[20] Initially, all of the chatbot's conversations were harmless banter with users, but soon, it turned into a reflection of the evil side of the internet. "Within 16 hours of Tay's release, the chatbot had tweeted more than 95,000 times, and a troubling percentage of her messages were abusive and offensive."[21] Most of these tweets were based on messages that were fed to the chatbot with the use of its "repeat after me" built-in function. Soon after, Twitter users began reporting the account, and Microsoft had little choice but to suspend the account. This demonstrates the fact that lack of human control and monitoring can lead to a supposedly "fun" experiment turning into a nightmare for companies, putting their reputation at stake.

Even though AI adoption can be a risky step to take for companies, if there are specific methods of mitigation put into place, the dangers can be combated. AI is a great tool to implement in a company's practices. Still, first and foremost, there is an imperative need for companies to do a thorough assessment of the potential impact on their organization's business model and existing strategies. This can be done by having it be a part of the innovation strategies in place. Stage gates processes within companies consist of "multiple reviews or gates"[22] that must be passed in order to consider an innovation as a successful venture to invest in. Another "gate" that should be added when innovations are in relevance to AI adoption is the ethical use of data to develop such a model. Companies need to consider the propriety of transactions when collecting and using data**.** This relates to asking questions such as "Is this legal, and right? Does it feel or look wrong?."[23] If and when companies begin to implement this practice of reflecting deeply on methods and strategies being made part of the organization's business model by asking the *right* questions, they will be able to promote integrity within business practices and respect an individual's right to privacy.

The risk of an algorithmic bias that can be caused by the developer of the algorithm and the quality of the data used can be combated in a couple of ways. One way to mitigate this risk is by understanding and addressing the potential for bias rather than denying the possibility of it occurring - this is half the task done. Once this step is accomplished, data scientists need to ensure that they conduct a deep evaluation of the data they are inputting into the models for training purposes. For this to happen successfully, there is a need for a higher level of transparency.[24] "Part of the move toward 'explainable AI' is to shine a light on how the data is being trained and how you're using which algorithms," said Jonathon Wright, chief technology evangelist at Keysight Technologies." This can go a long way in stopping the bias before it is instilled in AI models and further amplified. There is also a need to conduct background checks on the data scientists and developers involved in the making of AI models. This can be helpful in recruiting and assigning an ethical and rightly skilled right team to make an AI model that is as bias-free as possible. Deploying a framework consisting of ethical standards that are to be followed when developing AI models can be very useful in governing AI risks. Companies across the globe should bring in such a framework, as it will set the standard for the employees coming in, along with making them beware of the consequences for any breach of the code of conduct. This will help in re-instilling the idea of information integrity and reliability existing in AI models, which will lead to an enhanced trust factor between users and designers of AI models.

The quality of the datasets being used to train AI models can be improved by firstly increasing the size of the data being used in order to represent a larger population at least. It is not highly possible for results to be truly representative of every individual, but using data from several sources that are of higher quality, larger quantity, and one that's been tried and tested numerous times can help in reducing algorithmic bias that stems from poor quality of data.

One of the major risk exposures that an individual and a company face is one that stems from the lack of law enforcement in consumer protection policies. This can be changed if countries all around the globe begin to account for the technological developments of recent times. Current laws made to protect consumers were established long before AI sprung into existence and began to take over the world of technology. It is essential to have privacy laws be re-established and adjusted based on the current happenings. The European Union's GDPR (General Data Protection Regulation) does a much better job at updating its laws to the true need of the hour,[25] beneficial to consumers and individuals falling under that jurisdiction. According to GDPR, AI adopters have an obligation to provide *"meaningful information about the logic involved."*[26] The true essence of this requirement is to keep individuals informed of information you hold about them and how it is being used. So, if "you are going to use artificial intelligence to process someone's personal data, you normally need to tell them about it."[27] The United States and every other country should follow the path of the EU to reach a point of justice and higher protection for its citizens. Law enforcement is one of the most powerful ways to permanently fix issues in a manner that everyone must abide by, no matter the entity in question.

Talent strategy needs to be enhanced in firms to allow for a reduction in the chances of third-party and cyber risks. This could be accomplished by first building an in-house system within companies that can handle AI-related projects and escalations. AI is an evolving tool, and although it can be hard to keep up with, firms can at least begin to deploy appropriate resources (human and financial) to manage this within the internal system. This would mean allocating a budget for the handling, maintaining, and repairing of AI models. Cybersecurity professionals hold an integral role in assessing the impacts of AI in small and large scale deployment. Hence, the need to hire individuals with solid cybersecurity skills can prove to be of great benefit to companies that are aiming to expand their business through AI adoption. Training existing employees on how to use AI as a technological tool to fulfill requirements in innovations can also be very beneficial. Since these employees are already aware of how the business works, manipulating AI to aid the business can be made easier. Additionally, hiring new employees with versatility in skills to manage AI projects can be instrumental to a company's successful AI adoption. This is because they bring in a fresh and more tech-aware perspective, enabling the company to innovate in ways unimagined. With this, the need to outsource AI project management won't arise, which will, in turn, reduce the ways and means through which hackers can invade the internal systems and steal information.

The main question that is to be addressed through risk mitigation is how much human intervention is too much human intervention? Although human bias can very easily become a part of AI models, a lack of human oversight can lead to worse consequences. As Floridi mentioned, "Autonomous systems should always be subject to human oversight,"[28] and this translates to the fact that there is no need to "fear AI, but rather bring it into alignment with human principles and ethics."[29] It is a good idea for companies to create a balance between the level of human involvement needed depending on the situation. For example, human intervention is extremely important in constantly monitoring AI models deployed to account for processing errors, system failures, and algorithm manipulation made possible when models run at a larger scale and interact with a high volume of users. Having a clearly laid out plan of addressing concerns that AI models

may bring when made publicly available, along with developing a risk response strategy will genuinely provide a more effective use of a power tool like AI.

AI is an extremely powerful tool; however, this power is often abused and misused to benefit companies and political parties. This can be changed if we all take one step towards the responsibility we owe to society when using technology because no matter how small a change you make - it always counts in the larger picture.

SOURCES

[1] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Innovation Risks Lecture Slides.

[2] Team, D. editorial. (n.d.). Advantages of AI in Industrial Processes. Drew. Retrieved December 7, 2022, from https://blog.wearedrew.co/en/advantages-of-ai-in-industrial-processes

[3] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Emerging Risks Lecture Slides.

[4] Marr, B. (2021, July 13). What are the negative impacts of Artificial Intelligence (AI)? Bernard Marr. Retrieved December 7, 2022, from https://bernardmarr.com/what-are-the-negative-impacts-of-artificial-intelligence-ai/

[5] Ruhl, C. (2021, May 4). What is cognitive bias? Cognitive Bias: Definition & Examples - Simply Psychology. Retrieved December 7, 2022, from https://www.simplypsychology.org/cognitive-bias.html

[6] Ravi, R., & Ravi, A. R. (2022, May 13). AI gone wrong 5 biggest AI failures of all time. Jumpstart Magazine. Retrieved December 7, 2022, from https://www.jumpstartmag.com/ai-gone-wrong-5-biggest-ai-failures-of-all-time/

[7] Pratt, M. K. (2021, June 25). 5 ways AI bias hurts your business: TechTarget. Enterprise AI. Retrieved December 8, 2022, from https://www.techtarget.com/searchenterpriseai/feature/5-ways-AI-bias-hurts-your-business

[8] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Innovation Risks Lecture Slides.

[9] Ibid.

[10] Ibid.

[11] Johnston, L. (2020, April 20). Recent cases highlight growing conflict between AI and Data Privacy. Haynes Boone. Retrieved December 8, 2022, from https://www.haynesboone.com/news/publications/recent-cases-highlight-growing-conflict-between-ai-and-data-privacy

[12] Ibid.

13 Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Innovation Risks Lecture Slides.

[14] Breeden II, J. (2022, November 14). Expert analysis of Dangerous Artificial Intelligences in Government. Nextgov.com. Retrieved December 8, 2022, from https://www.nextgov.com/emerging-tech/2022/11/expert-analysis-dangerous-artificial-intelligences-government/379690/

[15] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Control Lecture Slides.

[16] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Types of Risk Lecture Slides.

[17] Ibid.

[18] Ibid.

[19] Schwartz, O. (2021, September 30). In 2016, Microsoft's racist chatbot revealed the dangers of online conversation. IEEE Spectrum. Retrieved December 8, 2022, from https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation

[20] Ibid.

[21] Ibid.

22 Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Innovation Risks Lecture Slides.

[23] Ibid.

[24] Apte, P. (2022, September 23). 5 ways to prevent Ai Bias. ITPro Today: IT News, How-Tos, Trends, Case Studies, Career Tips, More. Retrieved December 8, 2022, from https://www.itprotoday.com/artificial-intelligence/5-ways-prevent-ai-bias

[25] Apte, P. (2022, September 23). 5 ways to prevent Ai Bias. ITPro Today: IT News, How-Tos, Trends, Case Studies, Career Tips, More. Retrieved December 8, 2022, from https://www.itprotoday.com/artificial-intelligence/5-ways-prevent-ai-bias

[26] Church, P. (n.d.). Ai & THE GDPR: Regulating the minds of machines: DigiLinks. Linklaters. Retrieved December 8, 2022, from https://www.linklaters.com/en/insights/blogs/digilinks/ai-and-the-gdpr-regulating-the-minds-of-machines#:~:text='The%20GDPR%20also%20requires%20you,to%20tell%20them%20about%20it.

[27] Ibid.

[28] Searle, Annie. (2022, October). Info 312- Enterprise Risk Management (Fall 2022). Emerging Risks Lecture Slides.

[29] Ibid.