RAWATSAR P.G. COLLEGE
'Sanskriti Ka Badlta Swaroop Aur AI Ki Bhumika' (SBSAIB-2025)
DATE: 25 January 2025
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

Artificial Intelligence and its Impact on Society

Dr. Ranjeeta Garg, Asstt Professor, Department of Sociology, SMM Govt Girls College, Bhilwara ranjeeta.garg@gmail.com

Abstract

Artificial intelligence (AI) is a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyze data, make recommendations, and more. In other words, AI is the backbone of innovation in modern computing, unlocking value for individuals and businesses. It is a field of science concerned with building computers and machines that can reason, learn, and act in such a way that would normally require human intelligence or that involves data whose scale exceeds what humans can analyze.

AI has played a major role in the digitalization of society, as it has enabled us to collect, process, and analyze large amounts of data at a faster rate than ever before. This has led to the creation of new technologies, improved business processes, and greater efficiency in many industries. AI has the potential to revolutionize education also, offering personalized and individualized teaching, and improved learning outcomes. It provides numerous benefits such as reducing human errors, time saving capabilities, digital assistance, and unbiased decisions. However, the disadvantages include emotional intelligence, encouraging human laziness, and job Displacement.

While the benefits of AI are clear, there are also important ethical and societal implications that must be considered. Issues such as privacy, security, and job displacement are just a few of the challenges that come with the increasing use of AI. It is crucial that we address these concerns proactively and work to ensure that AI is used for the betterment of society. As AI continues to evolve and gain importance in our world, it is important that we continue to invest in its development and advancement.

Numerous significant societal sectors, such as healthcare, banking, and law enforcement, are already utilizing and being impacted by artificial intelligence. As AI capabilities advance, these applications will grow which might either seriously hurt society or have a very positive impact. In the end, the function of AI governance is to implement workable measures to reduce this danger of harm while permitting the advantages of AI innovation. This entails addressing difficult empirical problems regarding the present and possible risks and advantages of artificial intelligence (AI), evaluating effects that are frequently indirect and widely dispersed, and forecasting a very uncertain future.

It also necessitates considering the normative issue of what constitutes a beneficial application of AI in society, which is no less difficult. Even though several organizations may concur on broad concepts (such as privacy, equity, and autonomy) that AI applications should adhere to, difficulties arise when these principles are put into reality. For instance, it is obvious that AI systems must respect people's privacy, but most people would probably be ready to forgo some degree of privacy in order to create life-saving medical interventions. Research on these topics may and has advanced in spite of these obstacles. This paper's goal is to help comprehend both these advancements and the obstacles still facing us.

The present paper is an attempt to understand the ways in which AI is affecting society. An attempt is being made to understand the far reaching impacts of AI on society and human relations through a sociological perspective. The future of AI is bright and full of possibilities. As society continues to embrace this technology, it is crucial that we remain mindful of its impact and work to address the challenges that come with its evolution. By doing so, we can ensure that AI continues to play a positive role in our world, improving our lives and creating a better future for generations to come.

Keywords: AI, society, education, privacy, security

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

Introduction:

AI is already being used in and having an impact on a number of significant societal sectors, such as healthcare, banking, and law enforcement. We anticipate significant advancements in AI capabilities and their possible uses as long as funding for AI research is sustained, which will have even more significant societal effects. AI has the potential to significantly improve our understanding of the world and help us create fresh approaches to pressing issues like illness and climate change. But because AI systems are so powerful, they also run the risk of being deployed carelessly or with little regard for the immediate and long-term effects.

In the end, the function of AI governance is to implement workable measures to reduce this danger of harm while permitting the advantages of AI innovation. This calls for addressing difficult normative problems regarding what constitutes a positive application of AI in society as well as difficult empirical questions regarding the potential risks and advantages of AI. We must have a solid grasp of how AI is currently affecting society and how those effects are probably going to change in the future in order to appropriately weigh the dangers and rewards. It is difficult to evaluate even the immediate effects of a technology like artificial intelligence because they are probably extensively and unevenly dispersed throughout society. Additionally, it is challenging to assess how much of an impact AI systems have in comparison to other technology or societal shifts. Since it involves generating predictions about an unknown future, evaluating the possible effects of AI in the future—which is essential if we are to respond while consequences can still be shaped and harms have not yet occurred—is considerably more challenging.

The issue of what constitutes a beneficial use of AI in society is a complicated normative one. Numerous efforts have been made to define and reach consensus on high-level concepts, like autonomy, privacy, and justice that AI applications should adhere to (Jobin et al., 2019). Despite being a helpful initial step, putting these ideas into practice presents a number of difficulties. For instance, it seems obvious that using AI systems must respect people's privacy, but most people would probably be ready to forgo some degree of privacy in order to create life-saving medical treatments. Different cultures and groups will unavoidably hold differing opinions about the trade-offs we should make, and there might not be a simple solution or method for weighing opposing viewpoints.

In order to make decisions about AI that everyone will accept as valid, we must also create politically viable approaches to balance various viewpoints and values in real-world situations. Notwithstanding these difficulties, research can and has advanced our knowledge of AI's effects and the difficult normative issues they bring up. This chapter's goal is to help the reader comprehend both these advancements and the obstacles still facing us. Before moving on to some of the more alarming potential risks and sources of harm, we start by describing some of the advantages and opportunities AI has for society. Before offering some suggestions for AI governance today, we go over the various ethical and political issues that come up when attempting to strike a balance between these advantages and threats.

Advantages and prospects:

In the end, AI has promise for improving our understanding of the universe and problem-solving abilities beyond what humans could accomplish on their own. Three linked categories are used to discuss the possible advantages of AI: (1) extending and improving people's lives; (2) strengthening our collective capacity to address issues; and (3) fostering collaboration and moral advancement.

Increasing people's quality of life

AI can assist increase the effectiveness and quality of public services and products by customising them for a particular individual or situation. For instance, a number of businesses

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

have started utilising AI to provide individualised educational materials. This involves gathering information on students' performance and learning in order to better understand learning trends and individual learning requirements (Luan & Tsai, 2021). Similar to this, startups are starting to appear in the field of precision medicine, which uses AI to personalise healthcare by customising treatment based on unique patient characteristics. This field is still in its infancy but has great potential (K. B. Johnson et al., 2021).

AI has the potential to significantly advance our knowledge of illness and medicinal interventions. On a variety of specialised healthcare-related jobs, AI systems can currently perform better than human professionals. For instance, Google Health developed a model to predict the risk of breast cancer from mammograms, outperforming human radiologists in the process (McKinney et al., 2020). There is growing interest in using AI to improve drug discovery, for example, by more efficiently and quickly searching through and testing chemical compounds (Paul et al., 2021). In 2020, several startups in this field raised significant funds, and the first clinical trial of a drug created by AI started in Japan (Burki, 2020). The "protein folding" problem has seen significant advancements thanks to DeepMind's AI system AlphaFold, which could significantly enhance our capacity to treat illness (Jumper et al., 2021). According to Zhavoronkov et al. (2019), further advancements in AI for healthcare may potentially help us better comprehend and slow down the ageing process, leading to far longer lifespans than we currently experience.

Increasing our capacity as a society to address issues Through the modelling of the intricate systems underlying these issues, the advancement of the science underlying potential remedies, and the enhancement of the efficacy of policy interventions, artificial intelligence (AI) has the potential to assist in addressing many of the major issues facing our civilisation, including climate change and risks to global health.

Machine learning algorithms were used to characterise and predict the transmission patterns of the Zika (Jiang et al., 2018) and SARS-CoV-2 (Liu, 2020) outbreaks, demonstrating how AI can support early warning systems for threats like disease outbreaks and enable more timely planning and policymaking. In the future, it might be feasible to detect and contain such epidemics considerably sooner with improved data and more advanced methods. There has also been some preliminary discussion of the potential use of AI to spot early indicators of inequality and violence. For example, Musumba et al. (2021) estimate the likelihood of civil conflict in Sub-Saharan Africa using machine learning. Early conflict prevention intervention may become considerably simpler as a result. Resource management can be enhanced by AI-based models of complex systems, which may be especially crucial in reducing the consequences of climate change. For example, AI is starting to be used to forecast the grid's daily electricity need, increase efficiency, and figure out how best to distribute resources like car fleets to meet demand that is always fluctuating . In a similar vein, a greater comprehension of supply and demand in electrical grids might facilitate proactive management of an expanding variety of variable energy sources and lessen dependency on highly polluting plants (Rolnick et al., 2019). Numerous other issues, such as disaster response, could benefit from similar types of research. For instance, machine learning can be used to generate maps from aerial images and gather data from social media to guide relief efforts (Rolnick et al., 2019).

AI has the ability to improve science in important fields as well.

AI has the potential to enhance the scientific method in a variety of ways, including by assisting in the comprehension and visualisation of patterns in vast amounts of data or by carrying out more "routine" scientific research tasks like hypothesis generation, experimental design and analysis, literature search and summarisation, and more. The previously cited research on protein folding by DeepMind is an excellent illustration of how AI is already boosting science

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

in a significant field. By automating the laborious procedures involved in the discovery of new materials, artificial intelligence (AI) may accelerate advancements in fields such as materials science in the future. This might lead to the development of better materials for energy storage or harnessing, for instance (Rolnick et al., 2019).

AI can assist in determining the best existing remedies, as well as enhancing our comprehension of issues and developing the research required to address them. There is proof that by elucidating data uncertainties and enhancing current instruments for intervention design and evaluation, machine learning (ML) technologies can enhance policymaking (Rolnick et al., 2019). For example, Andini et al. (2018) demonstrate how a tax refund scheme may have been made more effective with the application of a straightforward machine learning method. AI might even be able to be used to create more capable organisations that could assist in solving a variety of issues. According to one theory, (human) participants could decide what goals an organisation should pursue and then let an AI system design the institution. This could provide innovative solutions to long-standing issues that people are unable to recognise. Facilitating ethical advancement and collaboration Most individuals would concur that most people now live in a better world than they did centuries ago.

This is partially because global living standards have increased as a result of economic and technical advancements. However, moral advancement is equally significant. For instance, because more and more individuals believe that sentient beings are deserving of care and moral concern, fewer humans and animals suffer nowadays. AI may speed up moral advancement, according to some theories (Boddington, 2021). For instance, it may play a "Socratic" role in assisting us in making better (moral) decisions for ourselves, drawing inspiration from Socratic philosophy's use of deliberative dialogue to help people form more moral judgements. Such systems could specifically aid in learning argumentation logic, enhancing conceptual clarity, bringing attention to one's own limits, and provide empirical backing for other viewpoints.

AI may also contribute to better group collaboration, which is perhaps the main factor in human accomplishment to date. Dafoe et al. (2020) list many ways AI could facilitate human collaboration: More sophisticated machine translation could help us get past real-world obstacles to greater international cooperation, such as increased trade and the potential for a more borderless world. AI tools could also assist groups in learning about the world together in ways that facilitate the development of cooperative strategies. Additionally, AI may be crucial in developing incentives for honest information exchange and investigating the realm of dispersed organisations that foster favourable cooperative behaviours.

Risks and harms:

Notwithstanding these numerous potential and actual advantages, we are already starting to observe negative effects from the usage of AI systems, which could worsen when more sophisticated systems are used on a larger scale.

This section will address five distinct ways that artificial intelligence (AI) could be harmful to people and society, each of which will outline current trends and effects of AI that point in that direction. We will also discuss what we might be particularly concerned about as AI systems become more capable and pervasive in society.

Increasing conflict's probability or intensity

Because AI makes it possible to create new and more deadly weapons, it may have an effect on how serious conflicts are. Lethal autonomous weapons (LAWs), which may have just been deployed in conflict for the first time, are of special concern. These are devices that can choose and engage targets without additional assistance from a human operator.There is compelling evidence that one form of deadly autonomous weapon, "armed fully autonomous drone swarms," is a weapon of mass destruction (WMD) (Kallenborn, 2020). This implies that they

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

would provide all of the risks associated with other WMDs, including the potential for geopolitical instability, terrorist attacks, and devastating wars between superpowers. Compared to the majority of other WMDs, they would also be more difficult to detect and safer to transport (Aguirre, 2020). Beyond LAWs, the use of AI in engineering or scientific research may pave the way for the creation of other incredibly potent weapons.
To make particularly ferocious biological viruses, for instance, it might be employed to determine the most hazardous genome sequences.
The possibility of an unintended or quick escalation in conflict may also rise as a result of the growing integration of AI into the defence and conflict domains. This is because it becomes more difficult to intervene to stop escalation when military decisions are increasingly automated (Deeks et al., 2018; J. Johnson, 2020). This is comparable to how financial sector algorithmic decision-making resulted in the 2010 "flash crash," where automated trading algorithms that lacked adequate monitoring produced a trillion-dollar stock market crash over around 36 minutes. Since there is no central authority to impose failsafe procedures, the repercussions in a conflict situation can be more worse than in the financial sector (J. Johnson, 2020).
AI may also change incentives in ways that increase the likelihood of conflict or cause it to worsen (Zwetsloot & Dafoe, 2019). For instance, by enhancing data collection and processing capabilities, AI could compromise second strike capabilities, which are essential to nuclear strategic stability. This would make it simpler to identify and eliminate previously secure nuclear launch facilities.

Increasing society's susceptibility to accidents or attacks:

The increasing integration of AI technologies into society may lead to new vulnerabilities that malicious actors could take advantage of. For example, by adding a tiny, undetectable disturbance to the image, researchers were able to trick an ML model that was trained to recognise traffic signals into identifying a "stop" sign as a "yield" sign (Papernot et al., 2017). Therefore, malicious actors may target an autonomous vehicle using this approach and change traffic signs with paint or stickers. Attacks of this nature may have increasingly disastrous outcomes when AI systems are used more extensively. For instance, adversarial attacks could endanger many lives if AI becomes more extensively incorporated into hospital diagnostic tools and our transportation networks.
In a similar vein, accidents may become more severe if increasingly powerful AI systems are used more widely.
In instance, integrating AI into vital infrastructure could increase efficiency, but it would also increase the risk of accidents on a much larger scale than is currently feasible. For instance, when autonomous vehicles proliferate, the failure of computer vision systems in severe weather or on roads may result in multiple vehicle crashes at once. There may be significant direct casualties as well as secondary consequences for supply chains and road networks. These kinds of malfunctions might potentially lead to the simultaneous failure of many vital systems, which at the most extreme could jeopardise the collapse of our entire civilisation if and when AI systems are developed to the point where they can manage sizable portions of society.
It would seem that ensuring that a human approves or makes the final choice might prevent these mishaps. But as AI capabilities like deep reinforcement learning (DRL) advance, we may be able to create more autonomous systems, and the market will probably push for their use. It is unclear how human oversight would function for such systems, particularly when their conclusions are too quick or difficult for people to understand (Whittlestone et al., 2021).
The competitive dynamics in AI development may make these threats worse. The development of AI is frequently described as a "race" between countries for technological superiority and

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

strategic benefit.

News outlets, the IT industry, and studies from government agencies like the Department of Defence and the U.S. Senate frequently use this framing. Actors working on AI may be more motivated to underinvest in the safety and security of their systems in order to remain ahead of the competition the more these competitive dynamics support AI development.

Increasing the concentration of power:

AI may significantly alter the way power is distributed in society, according to a number of connected phenomena. It appears likely that the advantages and disadvantages of AI will be distributed extremely unevenly throughout society in the absence of significant institutional change. AI systems are already discriminating against marginalised groups. For instance, it has been demonstrated that facial recognition software performs significantly worse for darker faces and an Amazon AI system devalued applications from job candidates whose resumes contained proof that they were female. Because marginalised populations typically have lower levels of technology literacy, they are also more vulnerable to the negative effects of AI, such as the spread of false information and deceptive advertising. Additionally, these populations are less likely to have the financial means to take advantage of AI advancements like tailored healthcare.

The development of AI is also increasing the wealth and power of already powerful and rich actors. Because they have access to the most data, computing power, and research talent, the companies with the largest market shares are able to create the most successful goods and services, which expands their market share and makes it simpler for them to keep accumulating data, compute, and talent. This strengthens the strong position that these tech businesses currently hold by creating a positive feedback loop. In a similar vein, wealthy nations that can invest more in AI development are probably going to see faster economic returns than emerging nations, which could cause the gap to increase. This could lead to a more severe concentration of power than we have ever seen, particularly if AI research results in faster economic growth than earlier technologies.

Furthermore, automation driven by AI has the potential to further widen the economic gap. It is inevitable that advancements in AI systems will enable the automation of a wider variety of tasks. More automation of manual labour occupations with lower salaries could result from advancements in reinforcement learning specifically, which could boost the dexterity and adaptability of robotic systems.

These individuals will have to retrain as a result of the automation of these jobs; even in the best scenario, their income would be temporarily disrupted. But not only manual work or low-paying employment are at jeopardy. Rapid automation of a variety of knowledge-based tasks, including as programming, creative writing, and journalism, may be facilitated by developments in language modeling. A large number of these

The highly skilled and sociable labour market, which is difficult to automate and already pays poorly, will be overrun by knowledge workers, further widening the income gap (Lee, 2018). Because algorithms are instantly distributable and infinitely replicable, unlike, say, computers or steam engines, and because highly effective venture capital funding is driving innovation, there is also reason to believe that changes in the availability of jobs due to AI may occur more quickly than previous waves of automation (Lee, 2018). This reduces the amount of time we have to get ready, such as by retraining people whose employment are most likely to be lost, and increases the likelihood that the effects on inequality will be more severe than ever before. AI advancements are also anticipated to provide governments and corporations greater influence over people's lives than in the past. Although only a few applications, including advertising and health, require extremely detailed data, the fact that existing AI systems need

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

a lot of data to learn from encourages businesses to gather more and more personal information from customers. People are becoming less able to give their consent or even understand how their data is being used, and powerful actors may be using this data collecting to monitor, influence, and even manage and control populations. In order to create a "search engine for faces," for instance, the startup ClearView AI harvested billions of photographs from Facebook, YouTube, Venmo, and millions of other websites. They then licensed the technology to more than 600 law enforcement agencies without any public scrutiny (Hill, 2020). Facial recognition technology is already being used in dangerous ways, such as monitoring Uighurs and other minority groups in China. The dual patterns of seemingly weakening privacy standards and growing AI use for population monitoring and manipulation are extremely worrisome.

Accordingly, AI can be used to target people and communities who are most likely to be open to it (e.g., through automated A/B testing) and to increase the production of convincing but inaccurate or misleading content online (e.g., through image, audio, and text synthesis models like BigGAN and GPT-3). More sophisticated versions would make it simpler for organisations to seek and maintain influence, for example by influencing elections or facilitating extremely effective propaganda, even though the negative effects of such tactics have so far been largely restricted. For instance, programs that "coach" their users to convince others of particular statements could be created using additional developments in language modeling. These tools have the potential to be utilised for social good, such as the New York Times chatbot that encourages users to get vaccinated against COVID-19, but they can also be used by self-interested parties to increase or maintain their power.

Undermining the capacity of society to resolve issues:

There may be wider adverse effects from the usage of AI in online information creation and distribution. Specifically, it has been proposed that social media corporations' use of AI to enhance their content recommendation engines is exacerbating online polarization.

In the future, our information ecology may be significantly impacted by the application of AI in information production or targeting. If numerous diverse groups utilise sophisticated persuasion techniques to promote a wide range of views, we may witness the globe breaking split into separate "epistemic communities," with limited opportunity for communication or information sharing. The growing personalisation of people's online experiences may lead to a similar situation: we might witness a continuation of the trend towards "filter bubbles" and "echo chambers," which are fuelled by content selection algorithms and which some claim are already occurring. Furthermore, greater understanding of these patterns in the creation and dissemination of information may make it more difficult for people to assess the reliability of any information source, which would lower information trust generally.

It would be far more difficult for humanity to make wise choices on significant issues under any of these scenarios, especially since there would be less faith in reliable multipartisan sources, which would thwart efforts at collaboration and group action. For instance, a lack of confidence in public health advice probably contributed to the vaccination and mask hesitation that worsened the effects of COVID-19 (Seger, 2021). We may envision a pandemic that is much more vicious, in which individuals take advantage of the situation to disseminate false information in order to achieve their personal objectives. This might result in risky behaviours, a much greater strain on healthcare systems, and far more disastrous consequences.

Letting AI systems take over the future:

We may also be concerned about humans losing control over crucial decisions if AI systems continue to advance in capability and start managing substantial portions of the economy. The only goal-directed process affecting the future at the moment is human endeavours to alter the planet. But this might alter with increased automated decision-making, which might mean that

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

humans lose part or all of their authority over the future. There are two presumptions that underlie this concern. First, AI systems will advance to the point where automating most economic tasks—from maintaining vital infrastructure to managing businesses—would not only be feasible but desired. Second, even with our greatest efforts, we might not fully comprehend these systems to ensure that they are in line with the desires of their operators.

To what extent are these assumptions believable? Taking the first into account, there is growing evidence that we could create AI systems this century that are just as competent as humans in a wide range of economically beneficial tasks. The development of human-level AI is the declared objective of two well funded organisations (DeepMind and OpenAI), as well as a respectable percentage of AI researchers. Massive resources are being invested in this field. The ability of AI to solve tasks for which they were not specifically trained and the performance gains that can be obtained by merely expanding the size of models, the datasets they are trained on, and the computational resources used for training them are just a few examples of the recent advances in AI that have defied expectations. For instance, GPT-3, OpenAI's most recent language model as of this writing, performs remarkably well on a variety of tasks that it was not specifically trained on, including creating functional code from descriptions in natural language, acting as a chatbot in specific situations, and serving as a creative prompt. A variety of commercial apps, such as GitHub Copilot, which helps programmers work more efficiently by proposing lines of code or complete functions, are rapidly being sparked by these capabilities. Simply increasing the size of earlier language models and training them with more data and processing power allowed for this advancement. There is strong evidence that this tendency will continue to produce increasingly potent systems without requiring "fundamental" advances in machine learning. Perhaps even more compelling is the second premise, which holds that sophisticated AI systems might not be entirely compatible with or intelligible to humans. AI systems are currently trained through "trial and error," whereby we look for a model that performs well on a particular target without necessarily understanding how a particular model generates the behaviour it does. As a result, we have no guarantee regarding the system's potential behaviour in novel settings.

One major worry is that AI systems may assist us in optimising for social metrics rather than what we truly value. For instance, we might use AI systems in law enforcement to help improve community security and safety, only to discover later that these systems are actually making people feel safer by reducing complaints and concealing information about shortcomings. It might be extremely expensive or even impossible to fix these kinds of issues if we don't recognise them until AI systems are used extensively in society. As was previously indicated, competitive pressures to use AI for financial advantage may increase the likelihood of this, leading actors to implement AI systems without enough guarantees that they are optimising for our desired outcomes.

Depending on the rate and direction of AI development, this could occur gradually or all at once. A much more gradual "takeover" of society by AI systems may be more likely, where humans don't fully realise they are losing control until society is nearly entirely dependent on AI systems and it is difficult or impossible for humans to regain control over decision-making. The most well-known versions of these concerns have focused on the possibility of a single misaligned AI system rapidly increasing in intelligence.

Political and ethical issues:

It is not very contentious to say that we should work to extend and improve people's lives and that we should prevent catastrophic accidents that take thousands of lives. But it's not always so simple to maximise AI's advantages while minimising its drawbacks. Sometimes the very item that entails risk is also what is required to permit a certain area of

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

advantage.

For instance, automating more and more aspects of the economy with AI has the potential to raise service quality and accelerate economic growth, both of which might significantly raise people's standard of living everywhere.

The economic benefits of this kind of advancement, however, as well as the negative effects of employment displacement, could be wildly disparate, resulting in a concentration of power and an unprecedented increase in social inequality. Research could advance on empirical questions on the processes that are most likely to worsen inequality. The possibility that the financial benefits of AI can be redistributed can also be increased by putting into practice some useful measures. When imagining what we want from AI in the future, however, there are still important moral decisions to be taken. For example, how should we weigh the possibilities of significant improvements in the average quality of life and societal advancement against the risk of drastically rising inequality? How much risk are we willing to accept if using AI in science has the potential to improve human health and lifespans but also runs the risk of producing harmful new technologies if not handled carefully and wisely? What should we do if delegating decision-making to AI systems could aid in the resolution of social issues that were previously unsolvable, but at the expense of diminished human autonomy and comprehension of the world?

There will be plenty of space for legitimate disagreement because these issues are normatively complicated. Today, people who value equality more than those who value collective well-being will wish to make different decisions. Values like human autonomy may be viewed quite differently in different cultures, younger individuals may be more willing to give up privacy than older generations, and those from nations with robust welfare states will probably be more worried about risks to equality.

How do we deal with these disagreements? In part, this is the domain of AI ethics research, which can help to illuminate important considerations and clearly outline arguments for different perspectives. However, we should not necessarily expect ethics research to provide all the answers, especially on the timeframe in which we need to make decisions about how AI is developed and used. We can also provide opportunities for debate and resolution, but in most cases it will be impossible to resolve disagreements entirely and use AI in ways everyone agrees with. We must therefore find ways to make choices about AI despite the existence of complex normative issues and disagreement on them.

According to some political scientists and philosophers, when reaching a consensus on final judgements is unachievable, we should instead concentrate on making sure the decision-making process is valid. Debates over public health ethics have also focused on decision-making processes rather than results; Daniels and Sabin (2008) contend that for decision-making processes to be seen as valid, they must be accessible to the public for review, revision, and appeal, among other reasons.

We do not currently have legitimate procedures for making decisions about how we develop and use AI in society. Many important decisions are being made in technology companies whose decisions are not open to public or even government scrutiny, meaning they have little accountability for the impacts of their decisions on society. For instance, despite being among the world's most influential algorithms, Facebook's and YouTube's content selection algorithms

are mostly opaque to those most impacted by them. The values and perspectives of individuals making important decisions have disproportionate influence over how "beneficial AI" is conceived of, while the perspectives of minority groups and less powerful nations have little influence.

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

Governance ramifications:

What steps should we take to try to guarantee that AI is created and applied in ways that are advantageous both now and in the future?
We propose three overarching goals for AI governance in the modern era.
In the end, AI governance ought to concentrate on determining and putting into place systems that maximise AI's advantages and minimise its drawbacks. But as we've covered in this chapter, there are two reasons why doing this might not always be simple. First, there are a lot of real and possible effects of AI that we do not yet fully comprehend to determine the advantages and disadvantages. Second, even where impacts are well understood,
tensions may arise, raising challenging ethical questions on which people with different values may disagree. AI governance therefore also needs to develop methods and processes to address these barriers: to improve our ability to assess and anticipate the impacts of AI; and to make decisions even in the face of normative uncertainty and disagreement. We conclude this chapter by making some concrete recommendations for AI governance work in each of these three categories.

Enabling benefits and mitigating harms:

In some cases, we might need to consider outright bans on specific applications of AI, if the application is likely to cause a level or type of harm deemed unacceptable. For example, there has been substantial momentum behind campaigns to ban lethal autonomous weapons (LAWs),3 and the European Commission's proposal for the first AI regulation includes a prohibition on the use of AI systems which engage in certain forms of manipulation, exploitation, indiscriminate surveillance, and social scoring. Another area where prohibitions may be appropriate is in the integration of AI systems into nuclear command and control, which could increase the risk of
accidental launch with catastrophic consequences, without proportional benefits.
However, effective bans on capabilities or applications can be challenging to enforce in practice. It can be difficult to achieve the widespread international agreement needed—for example, the US government have cited the fact that China is unlikely to prohibit LAWs as justification for not making the ban themselves. In other cases it may be difficult to delineate harmful applications clearly enough. In the case of the EU regulation, it is likely to be very difficult to clearly determine whether an AI system should be deemed as "manipulative" or "exploitative" in the ways stated, for example. If complete prohibitions are not practical, it might be possible to restrict access to potent features in order to lower the possibility of abuse. For instance, businesses may decide to restrict access to commercial items that have the potential to be misused or not release the complete code underlying particular features to stop bad actors from replicating them.
It will be essential to invest in socially beneficial applications, AI safety, and responsible AI research in order to move beyond harm prevention and reap the full benefits of AI. Many of the possible advantages we covered early in this chapter seem to have gone unnoticed. For instance, there could be much more focus on how AI might be used to improve moral reasoning, fight climate change, or foster intergroup collaboration. Working on these subjects may be hindered in part by the lack of strong incentives from industry (where financial incentives are not always in line with wide-ranging societal benefits) or academia (which frequently favours theoretical advancement over applications).

Enhancing our capacity to evaluate effects:

Before deciding what forms of governance are required, we frequently need to gain a deeper understanding of the possible effects of AI systems. There are several reasons why improved and standardised impact assessment procedures may be beneficial.

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 8.152

First, especially in safety-critical situations, we must define more precise guidelines and procedures for ensuring AI systems (also known as test, evaluation, validation, and verification, or TEVV) before they are put on the market. Much more work is required because there are currently no tried-and-true techniques for guaranteeing the behaviour of the majority of AI systems. It would also make it possible to identify and mitigate potential harms from unintended behaviour in advance, and to incentivise technical progress to make systems more robust and predictable.

Continual monitoring and stress-testing of systems will also be important, given it may not be possible to anticipate all possible failure modes or sources of attack in advance of deployment. Here it may be useful to build on approaches to 'read-teaming' in other fields including information and cyber security.

Additionally, we require more comprehensive methods to evaluate and predict the structural effects of AI systems. While assurance and stresstesting can assist in identifying potential harm from unexpected behaviours or attacks on AI systems, they are unable to detect situations in which a system operating as intended could nevertheless result in more extensive structural harms (such as polarising online discourse or altering incentives to increase the likelihood of conflict). This will likely require looking beyond existing impact assessment frameworks and drawing on broader perspectives and methodologies, including: social science and history, fields which study how large societal impacts may come about withoutanyone intending them (Zwetsloot & Dafoe, 2019); foresight processes for considering the future evolution of impacts (Government Office for Science, 2017); and participatory processes to enable a wider range of people to communicate harms and concerns.

Our capacity to foresee and get ready for new issues before they materialise might be enhanced by more methodical tracking of AI advancements (Whittlestone et al., 2021). Governments must be able to respond swiftly as technology develops and more AI systems are released onto the market, as these developments will present higher-stakes policy issues. The field of artificial intelligence (AI) is inherently generating a vast array of data, metrics, and measurements that might be included into a "early warning system" for emerging capabilities and applications that could significantly affect civilisation. Monitoring progress on widely studied benchmarks and assessment regimes in AI could enable AI governance communities to identify areas where new or more advanced applications of AI may be forthcoming.

Monitoring inputs into AI progress, such as computational costs, data, and funding, may also help to give a fuller picture of where societally-relevant progress is most likely to emerge. For example, early warning signs of recent progress in language models could have been identified via a combination of monitoring progress on key benchmarks in language modelling, and monitoring the large jumps in computational resources being used to train these models.

Conclusion

In this paper, an outline has been made about some of the possible ways AI could impact society into the future, both beneficial and harmful. Our aim has not been to predict the future, but to demonstrate that the possible impacts are wideranging, and that there are things we can do today to shape them. As well as intervening to enable specific benefits and mitigate harms, AI governance must develop more robust methods to assess and anticipate the impacts of AI, and better processes for making decisions about AI under uncertainty and disagreement.

References:

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., & Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in italy. Journal of Economic Behavior & Organization, 156, 86–102. https://doi.org/10.1016/j.jebo.2018.09.010

Boddington, P. (2021). AI and moral thinking: How can we live well with machines to enhance our moral agency? AI and Ethics, 1(2), 109–111. https://doi.org/10.1007/s43681-020-00017-0

Burki, T. (2020). A new paradigm for drug development. The Lancet Digital Health, 2(5), e226–e227. https://doi.org/ 10.1016/S2589-7500(20)30088-1

Daniels, N., & Sabin, J. E. (2008). Accountability for reasonableness: An update. BMJ (Clinical research ed.), 337, a1850. https://doi.org/10.1136/bmj.a1850

Jiang, D., Hao, M., Ding, F., Fu, J., & Li, M. (2018). Mapping the transmission risk of zika virus using machine learning models. Acta Tropica, 185, 391–399. https://doi.org/10.1016/j.actatropica.2018.06.021

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. Clinical and Translational Science, 14(1), 86–93. https://doi.org/https://doi.org/10.1111/cts.12884

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., ˇZ´ıdek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 1–11. https://doi.org/10.1038/s41586-021-03819-2

Kallenborn, Z. (2020). Are drone swarms weapons of mass destruction? U.S. Air Force Center for Strategic Deterrence Studies. Retrieved August 16, 2021, from https : / / media . defense . gov / 2020 / Jun / 29 / 2002331131/ - 1/ - 1/0/60DRONESWARMS-MONOGRAPH.PDF

Lee, K.-F. (2018). AI superpowers: China, silicon valley, and the new world order. Houghton Mifflin Harcourt

Liu, J. (2020, April 2). Deployment of health it in china's fight against the covid-19 pandemic [Imaging technology news]. Retrieved April 13, 2021, from https://www.itnonline.com/article/deployment-health-it-china%E2% 80%99s-fight-against-covid-19-pandemic

Luan, H., & Tsai, C.-C. (2021). A review of using machine learning approaches for precision education. Educational Technology & Society, 24(1), 250–266. Retrieved April 13, 2021, from https://www.jstor.org/stable/26977871

Musumba, M., Fatema, N., & Kibriya, S. (2021). Prevention is better than cure: Machine learning approach to conflict prediction in sub-saharan africa. Sustainability, 13(13), 7366. https://doi.org/10.3390/su13137366

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. arXiv:1602.02697 [cs]. Retrieved July 23, 2021, from http://arxiv.org/abs/1602. 02697

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N.,Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., . . . Bengio, Y. (2019). Tackling climate change with machine learning. arXiv:1906.05433 [cs, stat]. Retrieved April 12, 2021, from http://arxiv.org/abs/1906.05433

Seger, E. (2021). The greatest security threat of the post-truth age [BBC future]. Retrieved July 22, 2021, from https: //www.bbc.com/future/article/20210209-the-greatest-security-threat-of-the-post-truth-age

Whittlestone, J., & Clarke, S. (2022). AI challenges for society and ethics. In *The Oxford Handbook of AI Governance*. Oxford University Press.

Whittlestone, J., Arulkumaran, K., & Crosby, M. (2021). The societal implications of deep reinforcement learning. Journal of Artificial Intelligence Research, 70. https://doi.org/10.1613/jair.1.12360