Research Article

Towards a 'Responsible AI': Can India Take the Lead?

South Asia Economic Journal
21(1) 158–177, 2020
© 2020 Research and Information System for Developing Countries & Institute of Policy Studies of Sri Lanka
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/1391561420908728
journals.sagepub.com/home/sae
SAGE

Rajesh Chakrabarti[1]
Kaushiki Sanyal[2]

Abstract

The impact of artificial intelligence (AI) on every aspect of our lives is inevitable and already being felt in numerous ways. Countries are grappling with the opportunities and challenges that AI presents. Among the South Asian countries, India has taken a lead in promoting and regulating AI. However, it lags significantly behind countries such as China or the United States. This article explores India's AI ecosystem, the threats and challenges it faces, and the ethical issues it needs to consider. Finally, it examines the common concerns among South Asian nations and the possibility of coming together to promote and regulate AI in the region.

JEL: Z: Z0: Z000

Keywords

Artificial intelligence, machine learning, NITI Aayog, explainability, data protection, data localization

Introduction

Most believe that artificial intelligence (AI) will transform almost every aspect of our lives. However, there is considerable lack of clarity about the exact boundaries of AI in public discourse. Various definitions of AI as well as general usage of

[1] Jindal Global Business School, O. P. Jindal Global University, Sonipat, Haryana, India.
[2] Sunay Policy Advisory Pvt. Ltd., Gurgaon, Haryana, India.

Corresponding author:
Kaushiki Sanyal, CEO and Co-Founder, Sunay Policy Advisory Pvt. Ltd., 120/85, Silver Oaks Apartments, DLF City, Phase 1, Gurgaon, Haryana 122002, India.
E-mails: kaushiki.sanyal@gmail.com; kaushiki@sunayadvisory.com

the term have set the boundaries at different distances. Suffice it to say, AI is essentially about systems that can learn. It is a science and a set of computational technologies that are inspired by the ways people use their nervous systems and bodies to sense, learn, reason and take action. Thus, AI is demonstrated by a machine's ability to understand, think and act on a problem in the same way a human would in the same situation. There is no limit placed on the form the AI will take. AI can manifest itself invisibly on severs far away from the human eye working on advanced problems: it can take the form of a self-driving car, parts of a factory or even in the future some sort of advanced robotics.

The current debates have moved to questions of how, when and where the impact of AI will hit the hardest. Paradigmatic shift of this magnitude raises hope as well as concerns. AI can help eliminate disease and world poverty, improve productivity and enhance economies, but it can also take away jobs and throw millions of people into poverty. Although the exact nature of the changes that AI will bring to various sectors is not yet clear, some disruption to the workforce is virtually guaranteed. Furthermore, AI is having an impact across every industry—be it IT, manufacturing, retail, healthcare, financial services, education and media. It ranges from helping employees at transportation companies predict arrival times to predicting toxins in grains of food or helping scientists learn how to treat cancer more effectively. Businesses largely anticipate a positive impact on growth, productivity, innovation and in some cases job creation but challenges such as biases in algorithms, lack of data storage space and the need for massive skill upgradation remain (Marr, 2017). AI's challenges do not remain confined to unintentional or system-related threats. There are possibilities of AI being used with malicious intent consequences of which could be extremely dangerous for humanity. These include physical security risks where terrorists can repurpose commercial systems such as drones to deliver explosives or give low-skilled individuals the capability for high-skilled attacks with self-aiming, long range sniper rifles; political security risks where the government can use automated surveillance platforms to suppress dissent.

India is positioning itself to become the 'AI garage for emerging and developing economies' or a 'playground' for start-ups and enterprises globally to develop scalable solutions which can be easily implemented in other developing and emerging economies (NITI Aayog, 2018). Studies show that India has a workforce equipped with AI skills; its companies are among the early adopters of AI and it ranks third in research on AI. However, it lags behind in developing regulatory frameworks for personal data protection, standards of explainability, fairness appraisals, human-AI collaborations and liability frameworks.

This article provides an overview of India's AI ecosystem and its position in the world; the challenges it needs to address; the ethical issues and trade-offs it needs to consider while crafting a regulatory framework and the lessons that other South Asian countries can learn from India's experiences. The rest of the article is arranged as follows. In the section 'India's AI Ecosystem and Its Position in the World Research' we discuss Indian AI system and its position in world research. Risk and its challenges are addressed in the next section. After this, we discuss the necessity to craft a regularity framework for Indian AI. The role of South Asia countries in context to ‘Responsible AI’ is highlighted next and the final section concludes.

India’s AI Ecosystem and Its Position in the World Research

Research in AI across the globe has a long history of public funding with periodic ups and downs. However, the trend has moved towards private sector funding in recent years. Not surprisingly, the United States is the pioneer in AI research both in terms of public funding and breakthroughs. China is now rapidly catching up with the United States in both. Other countries like South Korea have turned to the public–private model for funding AI research.

For the United States, much of the funding came from DARPA’s Cyber Grand Challenge, a competition with prize money, and the European Union’s EU-FP7 technology funding programme. The BRAIN Initiative, created in 2013, is a 10-year, multibillion-dollar fund for AI research in the United States, while the EU’s Human Brain Project envisages spending 1 billion euros on AI over the next decade (Vempati, 2016).

How has research in AI fared in India? It lags significantly behind the United States and Europe and more recently China and South Korea. According to a 2012 report by Prof Deepak Khemani of IIT Madras, AI research in India has been limited to a handful of passionate researchers and a focus on only certain areas such as machine translation, natural language and text- and speech-related applications (Khemani, 2012). Unlike the United States where significant research on AI is undertaken by the Defense Advanced Research Projects Agency (DARPA), Indian AI research in defence is relatively limited. It is housed under the Centre for Artificial Intelligence and Robotics (CAIR), which is part of the Defence Research and Development Organization (DRDO). CAIR was established in 1986 and has worked on building integrated, networked information system, data mining tools, robotics and other AI-enabled products for the Indian military. It has had some successes but nowhere near the scale of other players.

A more recent 2018 study however shows that India ranks third in the world in terms of producing high quality research (computed as the number of citable documents in peer reviewed journals). Also, research has progressed to areas such as unsupervised learning, reinforcement learning, explainable AI, causal modelling and blockchain. The report also cited that researchers seem to receive adequate funding support from government, industry and universities.

However, the report identified challenges such as quality and quantity of students entering AI/ML research in India (it only has 50 to 75 principal researchers in the country), computing infrastructure, resources and administrative bottlenecks, lack of good quality labelled data sets and siloed research approach within universities (Itihaasa Research and Digital, 2018).

Investments

Although historically the government has been a key funder of AI, especially in the United States, investment by large corporations as well as funding from venture capital and private equity funds is growing fast. It is dominated by tech giants and digital native companies such as Alphabet (Google's parent company), Baidu, Apple, Facebook, Microsoft, IBM and Amazon who develop the inputs needed to enable AI applications—powerful computer hardware, increasingly sophisticated algorithmic models and vast inventory of data. These tech giants spent US$20 to 30 billion on AI in 2016—90 per cent on R&D and deployment and 10 per cent on AI acquisitions. VC and PE financing, grants and seed investments in AI start-ups also grew rapidly, albeit from a small base, to a combined total of less than 2 billion US$ in 2013 to over 6 billion US$ in 2017 (NASSCOM, 2018). Machine learning, as an enabling technology, received the largest share of both internal and external investment (Figure 1).

The USA is obviously the pioneer here. It boasts of the strongest ecosystem for AI in terms of funding, number of companies and global reach. About 40 per cent of all AI companies are based in the United States. Its leadership is a result of a mature, well-financed and thriving ecosystem in Silicon Valley and New York/ Boston metropolitan area. Over 16 governmental agencies support AI companies financially and politically (including DARPA, CIA and NSA), and it has leading universities (such as Stanford and MIT), as well as very strong corporate research facilities (such as Google DeepMind).

USA is followed, at a distance, by China, Israel and the United Kingdom. Approximately 13 per cent of all AI companies are based in China, 12 per cent in Israel and 8 per cent in the United Kingdom.

India lags quite a bit behind in terms of private sector investment in AI compared to the leading AI ecosystems globally, namely, the United States and China.

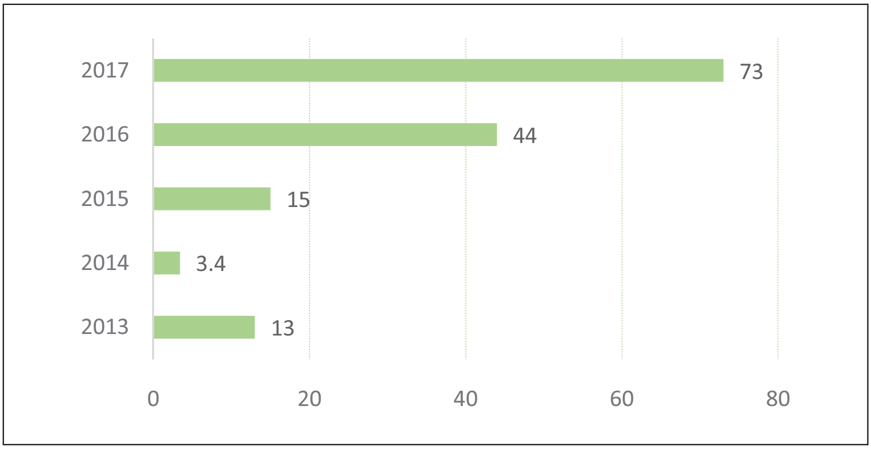

Figure 1. Funding for AI Companies in India (in US$ million)

Source: Artificial Intelligence Primer, NASSCOM, July 2018

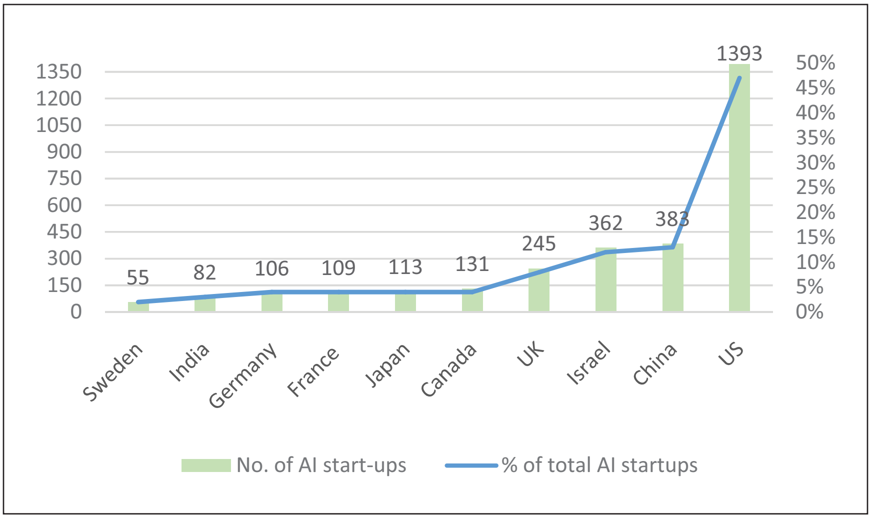

Figure 2. Top 10 Countries with Highest Number of AI Start-ups

Source: Global Artificial Intelligence Landscape, Asgard and Roland Berger, 14 May 2018. (https://www.rolandberger.com/publications/publication_pdf/roland_berger_ai_strategy_for_european_startups.pdf)

However, according to reports, there has been a jump in investment and deal activity around intelligent automation and artificial intelligence, machine learning and big data (Figure 2).

In 2018, start-ups raised an all-time high capital, registering a 368 per cent growth from 2017. In 2018, start-ups with operations in India and globally raised approximately US$ 529.52 million in funding rounds and this data includes start-ups with investment at varying stages of development, from pre-seed to well-funded companies. California and India based Automation Anywhere bagged the biggest cheque of US$ 300 million from SoftBank Vision Fund (Bhatia, 2018). The size of the AI market in India is estimated to grow to USD 89.8 billion in 2025 from US$ 3.2 billion in 2016 (Sachitanand, 2019).

India is also among the top 10 countries in the number of AI start-ups. In 2017, it had 82 AI start-ups (about 3 per cent of AI start-ups globally) based in the country. According to NASSCOM, the AI start-up pool is expanding rapidly at 54–58 per cent CAGR since 2013 (Figure 3).

The key segments are enterprise, marketplace, health-tech, ed-tech and fintech. Some of the emerging start-ups are: AnsweriQ (AI-enabled customer support ticket management solution), AskSid and Wysa (AI-based chatbot to understand user’s emotions). Among the mature start-ups, SigTuple (applies AI-powered analytics to visualize medical data), Flutura (combines AI and IoT to bring predictive analytics to manufacturing), Zendrive (AI based platform to monitor risky driving behaviour and triggering real time alerts), Lucep and Active.Ai (conversational banking platform to financial institutions) are key examples.

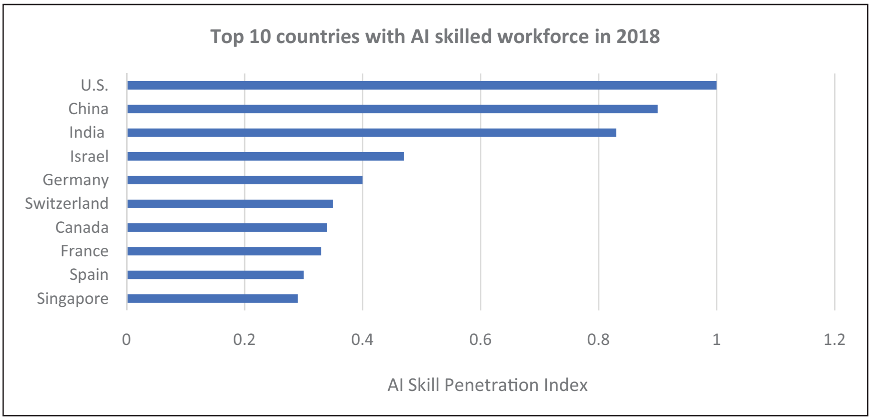

Figure 3. Countries with Highest Penetration of AI Skills among Its Workforce

Source: Statista (available at https://www.statista.com/statistics/947911/ai-skill-penetration-by-country/).

Note: The Index measured the changing nature of skills needed for different jobs. For example, an accountant focusing more on human-centric customer service tasks and less on routine tasks which have been automated or a data specialist using new programming and machine learning skills to more effectively target clients and ultimately generate revenue.

India is scrambling to make up for lost time which has led to some significant developments in the sector. The government has budgeted about Rs 3,073 crore in 2018, and there have been moves to focus on R&D and investment in AI. For example, Indian telcos Bharti Airtel and Reliance Jio have started research into AI through setting up AI labs. Global corporations such as NVIDIA, Microsoft and Google have set up R&D labs in India. Infosys, Wipro and other such IT giants have begun making equity investments in many AI-based start-ups (Anirudh VK, 2019).

Human Resources

India certainly has a size advantage. According to a report by LinkedIn (Perisic, 2018), India ranks among the top three countries in AI skills penetration after the United States and China, ahead of Israel and Germany (Figure 4).

However, among its 400 million odd labour forces, there is a large segment that is engaged in occupations such as agriculture, manufacturing and low-skilled white-collar jobs in the services sector. These people need to be re-skilled so that they do not lose their livelihood with the unfolding of the Fourth Industrial Revolution. In fact, a report published by the World Economic Forum (WEF) on the future of jobs states that in India 54 per cent of workers across 12 industries would need to be reskilled by 2022. While 35 per cent of the workers need at least six months of re-skilling, one in 10 would need over a year of training to be ready for the workplace of the future (World Economic Forum, 2018) (Figure 5).

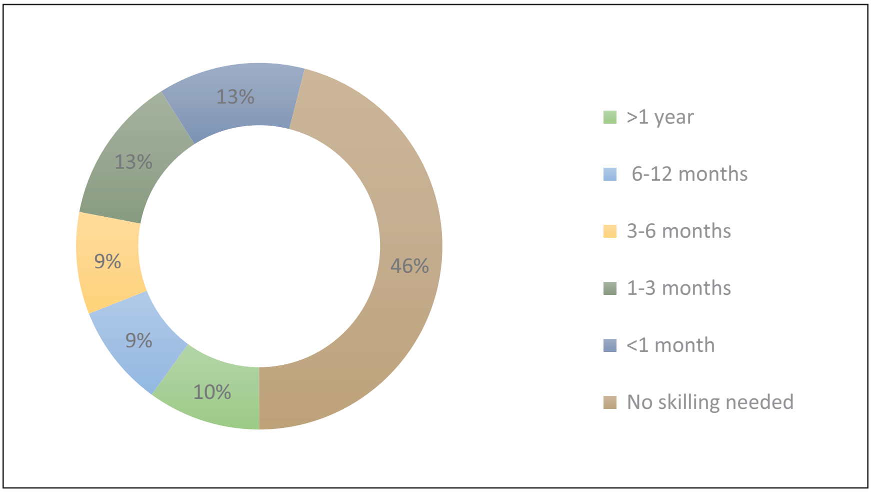

Figure 4. How Much Reskilling Does India Need?

Source: The Future of Jobs 2018, Insight Report, World Economic Forum, 2018 (available at http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf).

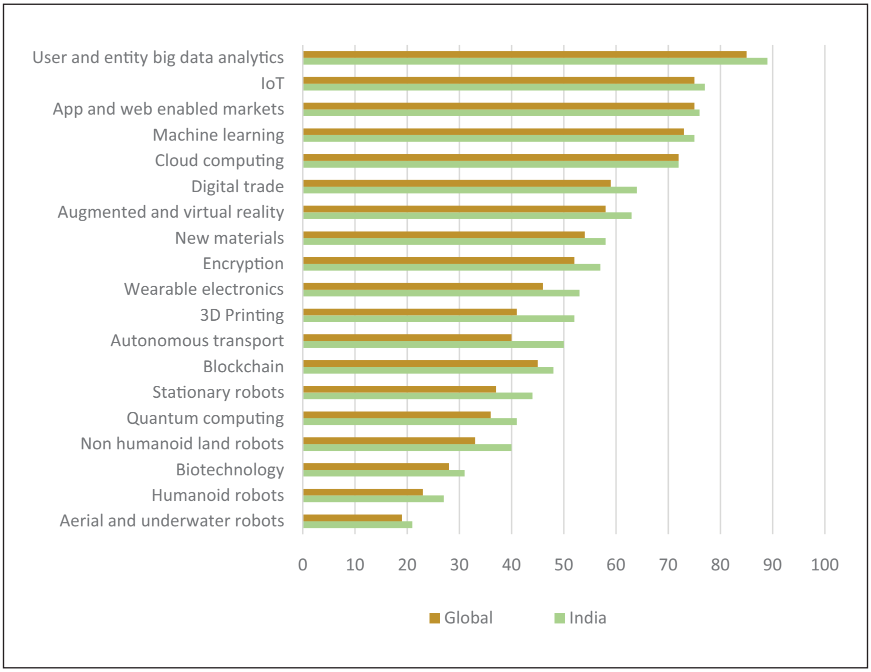

Figure 5. Technologies by Proportion of Companies Likely to Adopt Them by 2022 (Projected)

Source: Future of Jobs Survey 2018, World Economic Forum.

But there is hardly any clarity on the most unsettling question—its impact on the labour market. A two-year study from McKinsey Global Institute suggests that by 2030, intelligent agents and robots could eliminate as much as 30 per cent of the world's human labour. Depending upon various adoption scenarios, the report estimates that automation will displace between 400 and 800 million jobs by 2030, requiring as many as 375 million people to switch jobs categories (McKinsey Global Institute, 2017). Other studies predict that the impact of automation on jobs would probably be in the range of 14–54 per cent. While the estimation of automation may vary, it is clear that AI technology will change the business world in three aspects: automation, intelligence and creation. In most sectors, it will make some jobs, especially those requiring low skills, redundant while possibly increasing efficiency and creating new types of jobs (He & Guo, 2018).

What will the future of jobs look like in India? There are two key factors that pose a challenge (a) approximately 17 million entered the workforce year on year while only 5.5 million jobs are available; and (b) the speed and scale of the disruptions which is impacting the way we work and live. According to a report by EY, workforce in 2022 would look very different from today—9 per cent would be deployed in new jobs that do not exist today; 37 per cent would be deployed in jobs that require radically changed skill sets; and 54 per cent of jobs would fall under unchanged job category (Ernst & Young, 2017) (Table 1).

Adoption

Experts predict that AI will have applications across nearly every sector from education and healthcare to construction, retail and financial services. They however differ on the pace of AI adoption among Indian industries. A study by McKinsey suggests that the potential for adoption of AI technology and services in Indian industries is somewhat low because of inability to automate activities and digital absorption. But its human capital, innovation foundation, connectedness and labour market structure is within the global average

Table 1. A Sector-wise Snapshot of the Possible Deployment of the Workforce in 2022

	New Jobs That Do Not Exist Today (%)	Jobs with Radically Changed Skill Sets (%)	Jobs That May Be under Threat in 2017 (%)
IT	10–20	60–65	20–35
Automotive	5–10	50–55	10–15
Textiles	5–10	35–40	15–20
BFSI*	15–20	55–60	20–25
Retail	5–10	20–25	15–20

Source: Ernst and Young (2017; https://www.ey.com/Publication/vwLUAssets/ey-future-of-jobs-in-india/%24FILE/ey-future-of-jobs-in-india.pdf).

Note: *Banking, Financial Services and Insurance Sector.

(Bughin, Seong, Manyika, Chui, & Joshi, 2018). On the other hand, a survey conducted by the World Economic Forum among companies projects a more optimistic picture in terms of adoption of AI technology by 2022 in India. India is above the global average in almost every technology. Another survey, conducted by the Boston Consulting Group (BCG) in 2016 found that companies in the United States, China and India have taken an impressive lead in adoption of AI over their counterparts in Japan, France and Germany (Kupper, et al., 2018). Transportation and logistics, BFSI, automotive and technology companies are at the forefront of AI adoption.

World over adoption of AI outside the IT industry is slow. According to a global survey by MIT Sloan and BCG, only about one in five companies has incorporated AI in some offerings or processes. Only one in 20 companies has extensively incorporated AI in offerings or processes. Less than 39 per cent of all companies have an AI strategy in place. The largest companies—those with at least 100,000 employees—are the most likely to have an AI strategy, but only half have one (Ransbotham, Kiron, Gerbert, & Reeves, 2017).

Expectations from AI is running high but whether those expectations will be fulfilled depends on a variety of factors not limited to data mastery—flexibility in management and organizational practices, an AI strategy and an intuitive understanding of AI.

Risks and Challenges of AI

AI and cognitive solutions will not only change the business process but the entire business model that companies currently follow. The good news for India is that it is leading this revolution from the front and is going to be one of the fastest adopters of AI-based services. The Indian government has also woken up to the potential of AI and has started putting in place a strategy to scale up and allocate resources for research and training. However, there are many challenges that India would have to overcome before it can become the ‘AI garage’ for emerging economies. These include the lack of big data, digital infrastructure and highly trained man power.

It also has the potential to transform governance as it holds the key to changing millions of lives through dramatically improved delivery of public services, unprecedented efficiency in the design of law and order and regulatory monitoring systems. Governments, however, also have a critical role in not just harnessing AI wisely but also in developing it and regulating it to prepare society for adopting it gradually. Governments have a crucial responsibility in ensuring that AI applications create value for society, mitigate the adverse effects of job losses through safety nets and skill development and protect citizens from misuse of data. Failure on the part of policymakers to predict changes in society wrought by unhindered application of AI but private entities, especially in the jobs landscape could lead to political backlash.

We list some of the challenges and threats posed by AI.

Challenges

Scarcity of Big Data

The most powerful AI machines are the ones that are trained on supervised learning. This training requires labelled data—data that is organized to make it ingestible for machines to learn. However, the availability of well-labelled, feature-rich local data sets is extremely limited in India. A few government bodies make some data sets available, but they are limited in number and scope. For instance, the RBI maintains a database on the Indian economy, ISRO provides some data sets from its satellites via its mapping service Bhuvan, the Wildlife Institute of India provides some data sets that it tracks and maintains. Even the government's open data platform started in 2012 is sketchy. According to a study, critical data sets are not available on data.gov.in. Available data sets are often outdated, duplicated, incomplete, inadequately referenced and lack common terms used to describe the data. Top level metadata such as data collection methodology and a description of the variables are also either missing or incomplete. These shortcomings make it difficult to compare and analyse data sets properly. Organizations are trying to get around this issue by investing in design methodologies, trying to figure out how to make AI models learn despite the scarcity of labelled data. 'Transfer Learning', 'Unsupervised/Semi-Supervised Learning' and 'Active Learning' are just a few examples of the next-generation AI algorithms that can help resolve this.

Lack of Clean Data

For data to be used to train AI, it needs to be recorded in consistent, machine readable formats for accuracy and to ensure that it does not present the algorithms with unintended biases. This is a particularly big problem in India as a lot of its data is not digitized or in unstructured format. For example, India does not have a unified platform to access healthcare facilities at an affordable price. This basically means that there is no data repository available which can be hugely beneficial for both medical practitioners and patients as it becomes easier to track past medical history and also to provide better solutions by leveraging technologies such as AI and ML. Some companies are using international data-sets to overcome the problem. Chennai-based genomic intelligence company Kyvor Genomics uses AI models to develop its cancer therapy solution called CANLYTx. It is an AI-based system that involves a diagnostic test that identifies patients most likely to be helped or harmed by a new medication and based on that analysis, zeros in on targeted drug therapy. The company currently does not require local data sets as it is working with internationally available drugs for cancer treatment. Mohali-based agritech start-up Agnext developed a spectral analysis device that analyses the curcumin content in a particular turmeric harvest. Initially, they worked with world-wide crowd-sourced data-sets, but they started collecting data from labs across the country to build data sets.

Data Localization

The act of storing data on any device that is physically present within the borders of a specific country where the data was generated is known as data localization.

Free flow of digital data, especially data which could impact government operations or operations in a region, is restricted by some governments for security concerns. However, some experts oppose the move as it is seen as hindering the flexibility of the internet and adding to the cost for global companies who have to maintain multiple local data centres. Last year, India's Reserve Bank of India issued a circular mandating that payments-related data collected by payments providers must be stored only in India, setting 15 October 2018 as deadline for compliance. This covered not only card payment services by Visa and MasterCard but also of companies such as PayTm, WhatsApp and Google, which offer electronic or digital services. Countries such as China, Russia and Brazil also have strong data localization law, but China also has local data sets that can be used to train algorithms. India does not have that capability which could impact start-ups looking to attain global stature as reciprocal restrictions could be slapped by other countries. The United States, EU and Australia on the other hand allow for free flow of cross-border data to varying degrees.

Limited Technical Capacity

AI algorithms are usually very complex, often requiring thousands of calculations computed every second. As demand for more powerful processors increases, bottlenecks will start emerging, making it difficult for enterprises to adopt the technology. For start-ups and small and medium businesses, this would mean raising huge sums of capital to bring on board better processors and larger storage servers, which many would struggle to do. This trend also means that businesses will have a hard time securing data across multiple, non-relational databases that are constantly evolving.

Threats

A number of experts from various disciplines came together to contribute to a report that lays out the risks associated with AI being used with malicious intent. The report focuses on areas of AI that are available now or likely to be available within five years. Published in February 2018, the report warns that AI is ripe for exploitation by rogue states, criminals and terrorists (Brundage & Avin, 2018).

The threats outlined in the report are summarized below:

Threat to Physical Security

- **Terrorist repurposing of commercial AI systems:** Commercial systems can be used in harmful and unintended ways, such as using drones or autonomous vehicles to deliver explosives and cause crashes.
- **Endowing low-skill individuals with previously high-skill attack capabilities:** AI-enabled automation of high-skill capabilities—such as self-aiming, long-range sniper rifles—reduce the expertise required to execute certain kinds of attack.
- **Increased scale of attacks:** Human–machine can team up to use autonomous systems increasing the amount of damage that can be inflicted. For

example, one person launching an attack with many weaponized autonomous drones.

- **Swarming attacks:** Distributed networks of autonomous robotic systems, cooperating at machine speed, provide ubiquitous surveillance to monitor large areas and groups and execute rapid, coordinated attacks.
- **Attacks further removed in time and space:** Physical attacks are further removed from the actor initiating the attack as a result of autonomous operation, including in environments where remote communication with the system is not possible.

Threat to Political Security

- **State use of automated surveillance platforms to suppress dissent:** A nation's surveillance powers can be extended by automating image and audio processing, permitting the collection, processing, and exploitation of intelligence information at massive scales for myriad purposes, including the suppression of debate. For instance, Indian researchers from Cambridge University, India's National Institute of Technology, and the Indian Institute of Science presented a paper on a deep learning technique for facial image recognition to identify partially obscured faces. While the intended purpose is to nab criminals, it could be used by the state to target protesters and dissidents who conceal their faces at protests.
- **Fake news reports with realistic fabricated video and audio:** Creation of highly realistic videos showing inflammatory comments by influencers that they never actually made.
- **Automated, hyper-personalized disinformation campaigns:** Individuals can be targeted in swing districts with personalized messages in order to affect their voting behaviour.
- **Automating influence campaigns**: AI can be used to analyse social networks to identify key influencers, who can then be approached with offers or targeted with disinformation.
- **Denial-of-information attacks:** Bot-driven, large-scale information generation attacks are leveraged to swamp information channels with noise (false or merely distracting information), making it more difficult to acquire real information.
- **Manipulation of information availability:** Media platforms' content curation algorithms are used to drive users towards or away from certain content to manipulate user behaviour.

In addition to these threats which have a malicious intent, there are threats which are unintentional or system related such as algorithmic bias. It occurs when a computer system reflects the implicit values of the humans who created it. While generally the blame for bias in AI is put on the training data, the reality is that bias can creep in long before the data is collected as well as many other stages of the deep learning process—during the framing of the problem, collecting data and preparing the data. For example, biases creep in during hiring decisions as Amazon found out that its internal recruiting tool was dismissing female candidates because it was trained on historical hiring decisions which favoured males over females (Hao, 2019). In India, there is under-counting of crimes as the National Crime Records Bureau only records the ‘principal offence’ whenever a First Information Report is filed, which could mean that in a scenario where there is both rape and murder, murder could remain uncounted. This has huge implication for the future of predictive policing (Basu & Hickok, 2018).

Crafting a Regulatory Framework for India

Present Status

Currently, there is no comprehensive law that governs artificial intelligence in India. Over the years, the government has made piece-meal policies to protect certain facets related to AI, namely, **data privacy** and **data localization,** but there is not much discussion on possible regulatory issues beyond these. In fact, NITI Aayog’s AI Strategy Discussion Paper, published in 2018, in its section on ethics makes recommendation mostly on the issue of data privacy.

At present, the usage of personal data or information of citizens is regulated by the information technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011, under Section 43A of the Information Technology Act, 2000. The Rules define personal information of an individual as any information which may be used to identify them. They hold the body corporate (who is using the data) liable for compensating the individual, in case of any negligence in maintaining security standards while dealing with the data.

There were some attempts to frame data privacy laws between 2011 and 2015. A group of experts under Justice A.P. Shah had submitted a report on privacy in October 2012. The report proposed a framework for a Privacy Act in India which ‘must include privacy-related concerns around data protection on the internet and challenges emerging therefrom’. A draft Bill was also prepared post this report but did not reach the Parliament.

The introduction of India’s first biometric identity card for residents—Aadhaar—in 2012 brought the issue of data privacy to the fore-front. The Aadhaar data base captures biometric information (finger prints and retina scan) as well as basic demographic information (name, age, address, photo) and can be used to authenticate the identity of a person who wishes to avail a service provided by the government or a private sector organization. The concern with data privacy arises with the creation of a large data-base of residents and its use by third party service providers. Currently, the Aadhaar Act is silent on the powers of the UIDAI to take enforcement action against errant companies in the Aadhaar ecosystem. This includes companies wrongly insisting on Aadhaar numbers, those using Aadhaar numbers for unauthorized purposes and those leaking Aadhaar numbers, all of which have seen several instances in the recent past. Each of these can affect informational privacy and requires urgent redressal.

In this context, the right to privacy of citizens is the pillar on which India’s data protection regime has to be built. Since this right is not mentioned explicitly in the

Constitution, the matter has gone to court a number of times, the latest being in the *Puttaswamy* case (*Justice K.S. Puttaswamy (Retd.) v. Union of India*, 2017). Previous judgements on the right to privacy were in the context of the right to property and the surveillance powers of the state (Mittal, 2017).

In 2012, a petition was filed in the Supreme Court, challenging the constitutional validity of Aadhaar on the grounds that it violated an individual's right to privacy. The matter got referred from a three-judge bench to a five-judge bench and eventually to a nine-judge bench of the Supreme Court for an authoritative pronouncement on the status of the right to privacy. The bench gave its ruling on 24 August 2017 in what is referred to as the *Puttaswamy* case and affirmed the constitutional right to privacy. The Court also observed that 'informational privacy', or the privacy of personal data and facts, is an essential facet of the right to privacy (IndraStra, 2017).

The Supreme Court, however, clarified that like most other fundamental rights, the right to privacy is not an '*absolute right'.* Subject to the satisfaction of certain tests and benchmarks, a person's privacy interests can be overridden by competing state and individual interests. Thus, a violation of privacy in the context of an arbitrary State action would attract a '*reasonableness'* enquiry under Article 14. Similarly, privacy invasions that implicate Article 19 freedoms would have to fall under the specified restrictions under this constitutional provision such as public order, obscenity, etc.; and the intrusion into life or personal liberty under Article 21, which forms the '*bedrock of the privacy guarantee*', would have to be just, fair and reasonable. Lastly, the court mentioned a fourth test for deciding whether privacy was breached—the '*highest standard of scrutiny' which* can be justified only in case of a '*compelling state interest'.*

The government, on its part, constituted a Committee of Experts to deliberate on a data protection framework for India in July 2017 under the chairmanship of Justice B.N. Srikrishna, former judge of the Supreme Court. The committee's mandate was to develop a framework to '*ensure the growth of the digital economy while keeping personal data of citizens secure and protected*' (Srikrishna, 2018). The committee submitted its report in July 2018 along with a draft Personal Data Protection Bill, 2018.

The union government is considering introducing the Personal Data Protection Bill, a draft of which was prepared in 2018, but the formal introduction of which was held back till after the 2019 general elections.

The other related area of the digital ecosystem that has received attention in India is 'data localization'. Over the past year, the government has drafted and introduced multiple policy instruments which dictate that certain types of data must be stored in servers located physically within the territory of India. Presently, India has four sectoral policies that deal with localization requirements based on type of data, for sectors including banking, telecom and health: the RBI Notification on 'Storage of Payment System Data'; the FDI Policy 2017; the Unified Access License; the Companies Act, 2013 and its Rules; the IRDAI (Outsourcing of Activities by Indian Insurers) Regulations, 2017; and the National M2M Roadmap.

At the same time, 2017 and 2018 have seen three separate proposals for comprehensive and sectoral localization requirements based on the type of data across sectors including the draft Personal Data Protection Bill 2018, draft e-commerce policy and the draft e-pharmacy regulations. The policies reflect objectives such as enabling innovation, improving cyber security and privacy, enhancing national security and protecting against foreign surveillance (Basu, Hickok, & Singh Chawla, 2019).

While there is merit in data localization for reasons of preserving data sovereignty, there are risks that should be considered. These include impact on India's trade relationship, security risks (storing data in multiple physical centres increases the physical exposure to exploitation by individuals physically obtaining data or accessing the data remotely), economic fallout (it would increase entry barriers and compliance cost for foreign service providers).

India's AI regulations are still at a nascent stage and needs to evolve significantly on issues related to human-AI collaboration, general liability frameworks, fairness appraisals, explicability standards and safety considerations. Government needs to collaborate with relevant stakeholders especially AI practitioners to evolve standards and guidelines to ensure that AI technology remains socially beneficial while contributing to the economic growth of the country.

Ethical Issues to Consider

Fairness

Machine learning can significantly improve accuracy relative to most traditional decision-making processes. Its value can come from better resource allocation decisions as well as improving efficiency and effectiveness of government programmes. But AI algorithms and data sets can also reflect, reinforce or reduce unfair biases.

Thus, the government needs to ensure that a clear baseline accuracy for decision-making exists before implementing an algorithm whether based on historical human decisions, rudimentary scoring or criteria-based approaches that were being used. In India, AI could possibly be used to accurately allocate resources for welfare programmes or target beneficiaries of welfare programmes. However, given the huge inclusion and exclusion errors in the existing list of beneficiaries, there needs to be a much greater effort in laying down transparent, easy-to-implement standards of inclusion. Also, a cost–benefit analysis of the various regulatory tools for tackling the problem of fairness—self-certification, certification by a self-regulatory body, discrimination impact assessments and investigation by the privacy regulator—is required before developing a regulatory framework.

Accountability and Remediability

Algorithms make decisions with far-reaching impact on the lives of humans, especially the most socio-economically disadvantaged. But they can also yield unfair and discriminatory outcomes. Therefore, algorithmic systems need to be held accountable in the interest of justice and fairness. But how does one hold algorithms accountable? Accountability may be achieved by human audits, impact assessment or via governance through policy or regulation. Governance through 'human in the loop', where certain decisions identified as high-risk require vetting by a human, is a possible model.

There also needs to be redressal mechanisms if bias or inaccuracy is proven. Unfair systems should either be withdrawn or modified within a specified time frame. Furthermore, if the AI is involved in the commission of a crime or the violation of human rights, it needs to be held accountable. However, who should be held accountable and how should the accountability be enforced are questions that need to be given deeper considerations.

Transparency and Explainability

A lot of the ethical concerns around AI stem from its inherent 'black box' behaviour. This is partly because companies do not want to share the 'secret sauce' that makes their model click, and partly because so much of the learning in machine learning is locked in large complex math operations. There is now research on interpretability (opening up the black box or the process) and explainability (understanding the decision), which is essential for developing regulatory standards for both. Given the biases of caste, class and gender prevalent in our police force as well as administrative machinery, India needs to adequately understand these processes and develop standards before deploying AI for identifying beneficiaries or in law and order.

Security and Safety

When AI is deployed in critical areas with potential for greater harm such as healthcare and autonomous transportation, regulations need to include appropriate testing and quality assurance standards. For example, how do we regulate self-driving cars? Should they be given a licence after a test drive? Who should be held accountable if there is a misdiagnosis based on faulty algorithm? Also, there needs to be adequate protection for whistle-blowers who report privacy breaches and vulnerabilities.

Taking the Lead in 'Responsible AI': Lessons for South Asian Economies

The government is keen to position India as a leader among developing economies on AI related issues. However, it has yet to frame comprehensive regulations to ensure that Indian citizens are not used as guinea pigs for technologies whose effects are unknown. Other South Asian countries, keen to get on the AI bandwagon, should be cautious about allowing themselves to become testing ground for untried technologies.

Given the lack of norms surrounding the use of AI systems, India and other South Asian economies could take the lead in developing standards and protocols in the use of AI in both civil and military spheres. India has made a start in this direction by forming various expert committees to develop regulations, and other South Asian economies could follow the same path. Given the similarities in the problems facing these countries, pooling the collective wisdom of experts in these countries would help in developing a robust regulatory framework.

It is a fact that AI will impact the labour market massively. Given that India has a comparative advantage among the South Asian countries in terms of its educational facilities, there could be more collaborations between India and South Asian countries to promote research through creation of research hubs and centres of excellence in key universities where researchers from various South Asian countries are welcome to use the facilities as well as incubation centres that are open to any citizen of south Asian countries.

There could also be forums at various levels for exchanging information about the challenges faced in the deployment of AI in various sectors such as among schools, hospitals and the police force. This could help in problem solving as well as evolving best practices in regulation of AI.

Concluding Remarks

As both the promise and risks of AI are likely to impact across countries, industries and social classes, governments need to be proactive in not only harnessing AI technology for economic growth but also put in place regulations to ensure that citizens are protected from the threats posed by AI. However, given the early stage of AI development, it is important to focus on laws and norms that retain flexibility as new possibilities and problems emerge. This is particularly crucial given that AI is multi-purpose in nature. It is also imperative that countries cooperate and collaborate with each other at various levels—government, academia, civil society and corporates—to develop regulatory frameworks that address the challenges and risks posed by AI. The spill-over effects of contradictory regulations across countries and working in silos could be immense given the scale of threats that AI can pose on the security and sovereignty of a country.

Every regulation that is developed needs to debate the trade-offs between many factors: how stringent should standards of explainability be? What should be the definition of fairness as there are conflicting definitions? How should safety problems be addressed? But in setting benchmarks, it is important to factor in the opportunity cost of not using an AI solution when one is available; and to determine at what levels of relative safety performance AI solutions should be used to supplement or replace existing human ones. AI systems can make mistakes, but so do people, and in some contexts, AI may be safer than alternatives without AI, even if it is not fail-proof.

Finally, we need to keep in mind that AI is a tool that can be applied with good or ill-intent. We need to frame regulations in such a way that would minimize harm without Impacting technological breakthroughs.

Acknowledgements

The article was presented at the 12th South Asia Economic Summit (SAES) in Colombo, Sri Lanka, 26–27 September, 2019. Authors are grateful to Prabir De, Professor, RIS, India; Selim Raihan, Executive Director, South Asian Network of Economic Modelling, Bangladesh; Puspa Sharma, Deputy Executive, SWATEE, Nepal; and conference participants for their valuable comments. Usual disclaimers apply.

Box 1. Recommendations Related to Ethics and Privacy in NITI Aayog's Paper

- **Establish a data protection framework with legal backing:** This should be based on the seven core principles of data protection and privacy – informed consent, technology agnosticism, data controller accountability, data minimization, holistic application, deterrent penalties and structured enforcement.
- **Establish sectoral regulatory frameworks:** In addition to a central privacy protection law, sectoral regulatory frameworks are needed to protect user privacy and security. Japan and Germany have developed new frameworks applicable to specific AI issues such as regulating next generation robots and self-driving cars respectively.
- **Benchmark national data protection and privacy laws with international standards:** EU's General Data Protection Regulation (GDPR) guidelines, which have been enforced in May 2018, and the French laws could act as guidelines for India's privacy protection regime.
- **Encourage AI developers to adhere to international standards:** AI practitioners from across the world have acknowledged the need to frame standards for AI and many have set out guidelines to be followed. Indian enterprises need to step up too.
- **Encourage self-regulation**: Data Privacy Impact Assessment Tools can be used by AI developers and enterprises to manage privacy risks.
- **Invest and collaborate in privacy preserving AI research:** New mathematical models for preserving privacy are undergoing R&D and India should collaborate on areas of research such as Differential Privacy, Privacy by Design, Safety-Critical AI and Multi-Party Computations which enable protection of privacy despite data sharing at a wide scale.
- **Spread awareness: The Supreme Court of India has termed p**rivacy as a fundamental right but this right can be protected not only by laws enforced by state but making the citizens aware of their rights and how they can protect them.

Source: National Strategy for Artificial Intelligence: #AIForAll, Discussion Paper, NITI Aayog, June 2018.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

Basu, A., & Hickok, E. (2018). *AI in the governance sector in India.* The Centre for Internet and Society. Retrieved from https://cis-india.org/internet-governance/ai-and-governance-case-study-pdf

Basu, A., Hickok, E., & Singh Chawla, A. (2019, March 19). *The localisation gambit: Unpacking policy measures for sovereign control of data in India.* The Centre for Internet and Society. Retrieved fromhttps://cis-india.org/internet-governance/resources/the-localisation-gambit.pdf

Bhatia, R. (2019, January 15). *Indian AI startup funding 2018: Total global investment in India touched USD$529.52 million.* Analytics India Magazine. Retrieved from https://www.analyticsindiamag.com/indian-ai-startup-funding-total-global-investment-in-india-touched-usd-529-52-million/

Brundage, M., & Avin, S. (2018, February). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation.* Future of Humanity Institute, University of Oxford; Centre for the Study of Existential Risk, University of Cambridge; Center for New American Security; Electronic Frontier Foundation; OpenAI. Retrieved from https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/MaliciousUseofAI.pdf?ver=1553030594217

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy.* McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-ec

Ernst and Young. (2017). *Future of jobs in India: A 2022 perspective.* Kolkata: Ernst and Young. Retrieved from https://www.ey.com/Publication/vwLUAssets/ey-future-of-jobs-in-india/%24FILE/ey-future-of-jobs-in-india.pdf

Hao, K. (2019, February 4). This is how AI bias really happens: And why it's so hard to fix. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

He, D., & Guo, V. (2018, September 14). *4 ways AI will impact the financial job market.* World Economic Forum. Retrieved from https://www.weforum.org/agenda/2018/09/4-ways-ai-artificial-intelligence-impact-financial-job-market/

IndraStra. (2017, November 18). *An analysis of Puttaswamy: The Supreme Court's privacy verdict.* Retrieved from https://medium.com/indrastra/an-analysis-of-puttaswamy-the-supreme-courts-privacy-verdict-53d97d0b3fc6

Itihaasa Research and Digital. (2018). *Landscape of artificial intelligence/machine learning research in India.* (2018). Retrieved from http://www.itihaasa.com/pdf/itihaasa_AI_Research_Report.pdf

Justice K.S. Puttaswamy (Retd.) v. Union of India, Writ Petition (Civil) No. 494 of 2012 (Supreme Court August 24, 2017). Retrieved from https://www.sci.gov.in/supremecourt/2012/35071/35071_2012_Judgement_24-Aug-2017.pdf

Justice Srikrishna, B. (2018). *A free and fair digital economy: Protecting privacy, empowering Indians.* New Delhi: Ministry of Electronics & Information Technology, Government of India. Retrieved from https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report-comp.pdf

Khemani, D. (2012). A perspective on AI research in India. *AI Magazine, 33*(1), 96–98. doi:10.1609/aimag.v33i1.2356.

Kupper, D., Lorenz, M., Kuhlman, K., Bouffault, O., Lim, Y., Van Wyck, J., … Schlageter, J. (2018). *AI in the factory of the future: The ghost in the machine.* The Boston Consulting Group. Retrieved from https://www.bcg.com/en-in/publications/2018/artificial-intelligence-factory-future.aspx

Marr, B. (2017, July 13). *The biggest challenges facing artificial intelligence (AI) in business and society*. Retrieved from https://www.forbes.com/sites/bernardmarr/2017/07/13/the-biggest-challenges-facing-artificial-intelligence-ai-in-business-and-society/#e35448f2aec7

McKinsey Global Institute. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. McKinsey & Company. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx

Mittal, P. (2017, July 19). Is privacy a fundamental right? Two cases that Supreme Court will look at. *Livemint*. Retrieved from https://www.livemint.com/Politics/7oHGx6UJfLD0uIDXFwV9CL/Is-privacy-a-fundamental-right-Two-cases-that-Supreme-Court.html

NASSCOM. (2018). *Artificial intelligence primer*. New Delhi. Retrieved from https://community.nasscom.in/wp-content/uploads/attachment/nasscom-ai-primer-2018.pdf

NITI Aayog. (2018). *National strategy for artificial intelligence: #AIforAll*. New Delhi: Government of India. Retrieved from https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

Perisic, I. (2018, September 17). *How artificial intelligence is already impacting today's jobs*. LinkedIn. Retrieved from https://economicgraph.linkedin.com/blog/how-artificial-intelligence-is-already-impacting-todays-jobs

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). *Reshaping business with artificial intelligence*. MIT Sloan Management Review, The Boston Consulting Group. Retrieved from https://www.bcg.com/Images/Reshaping%20Business%20with%20Artificial%20Intelligence_tcm21-177882.pdf

Sachitanand, R. (2019, February 10). *Here's why Indian companies are betting big on AI*. Retrieved from Economic Times: https://economictimes.indiatimes.com/tech/internet/heres-why-indian-companies-are-betting-big-on-ai/articleshow/67919349.cms

Vempati, S. (2016, August 11). *India and the artificial intelligence revolution*. Retrieved from Carnegie India: https://carnegieindia.org/2016/08/11/india-and-artificial-intelligence-revolution-pub-64299

VK, A. (2019, January 10). *Is India finally catching up to China's state powered AI dreams?* Analytics India Magazine. Retrieved from https://www.analyticsindiamag.com/is-india-finally-catching-up-to-chinas-state-powered-ai-dreams

World Economic Forum. (2018). *The future of jobs 2018*. Insight Report, Centre for the New Economy and Society, Switzerland. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf