NITI Aayog

NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE #AIFORALL

JUNE 2018

This page has been intentionally left blank

Acknowledgments

In writing this Report, Arnab Kumar, Punit Shukla, Aalekh Sharan and Tanay Mahindru of NITI Aayog have made valuable contributions.

Inputs were also provided by Dr. Avik Sarkar, Dr. Ashish Nayan and Kartikeya Asthana of NITI Aayog.

The valuable contributions of Mr. P. Anandan and his team from Wadhwani Institute for AI, Dr. Manish Gupta of Videoken, Prof. Ramesh Raskar of MIT Media Labs, nVIDIA, Intel, IBM, NASSCOM, McKinsey and Accenture are also thankfully acknowledged. Special contribution of Accenture in the production of this report is also acknowledged.

Anna Roy
Advisor (Industry)
NITI Aayog

Contents

INTRODUCTION 5

EXECUTIVE SUMMARY 7

WHAT IS ARTIFICIAL INTELLIGENCE? 12

GLOBAL DEVELOPMENTS IN ARTIFICIAL INTELLIGENCE 16

ARTIFICIAL INTELLIGENCE AND INDIA 18

FOCUS AREAS FOR AI INTERVENTION 24
- Healthcare 24
- Agriculture 30
- Education 35
- Smart Cities and Infrastructure 39
- Smart Mobility and Transportation 41

KEY CHALLENGES TO ADOPTION OF AI IN INDIA 46

WAY FORWARD TO HARNESS THE POWER OF AI 48

RESEARCH 50

SKILLING FOR THE AI AGE 64

ACCELERATING ADOPTION 71

ETHICS, PRIVACY, SECURITY AND ARTIFICIAL INTELLIGENCE 85

ACTIONS FOR THE GOVERNMENT 91

APPENDIX I: ARTIFICIAL INTELLIGENCE EXPLAINED 96

APPENDIX II: GLOBAL COUNTRY STRATEGY REVIEW 100

APPENDIX III: DATA ECOSYSTEM 110

Introduction

#AIforAll: Technology Leadership for Inclusive Growth

Artificial Intelligence (AI) is poised to disrupt our world. With intelligent machines enabling high-level cognitive processes like thinking, perceiving, learning, problem solving and decision making, coupled with advances in data collection and aggregation, analytics and computer processing power, AI presents opportunities to complement and supplement human intelligence and enrich the way people live and work.

India, being the fastest growing economy with the second largest population in the world, has a significant stake in the AI revolution. Recognising AI's potential to transform economies and the need for India to strategise its approach, Hon'ble Finance Minister, in his budget speech for 2018 – 2019, mandated NITI Aayog to establish the National Program on AI, with a view to guiding the research and development in new and emerging technologies. In pursuance of the above, NITI Aayog has adopted a three-pronged approach – undertaking exploratory proof-of-concept AI projects in various areas, crafting a national strategy for building a vibrant AI ecosystem in India and collaborating with various experts and stakeholders. Since the start of this year, NITI Aayog has partnered with several leading AI technology players to implement AI projects in critical areas such as agriculture and health. Learnings from these projects, under various stages of implementation, as well as our engagement with some of the leading institutions and experts have given a better perspective to our task of crafting the national strategy for AI, which is the focus of this discussion paper.

This strategy document is premised on the proposition that India, given its strengths and characteristics, has the potential to position itself among leaders on the global AI map – with a unique brand of **#AIforAll**. The approach in this paper focuses on how India can leverage the transformative technologies to ensure social and inclusive growth in line with the development philosophy of the government. In addition, India should strive to replicate these solutions in other similarly placed developing countries.

#AIforAll will aim at enhancing and empowering human capabilities to address the challenges of access, affordability, shortage and inconsistency of skilled expertise; effective implementation of AI initiatives to evolve scalable solutions for emerging economies; and endeavors to tackle some of the global challenges from AI's perspective, be it application, research, development, technology, or responsible AI. #AIforAll will focus on harnessing collaborations and partnerships, and aspires to ensure prosperity for all. Thus, #AIforAll means technology leadership in AI for achieving the greater good.

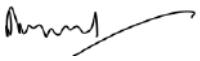
While evolving the national strategy for AI, the underlying thrust was to identify applications with maximum social impact, a willingness to learn from the best of the world when it comes to the recent technology advancements in AI, and leveraging approaches that democratize access to and further development of AI.

From an applications perspective, the approach is to identify sectors that may have the potential of greatest externalities while adopting AI solutions, and hence require the government to play a leading role in developing the implementation roadmap for AI. For example, the agriculture sector in India, which forms the bedrock of India's economy, needs multi-layered technology infusion and coordination amongst several stakeholders. Efforts from private sector may neither be financially optimal nor efficient on a standalone basis, and hence sustained government intervention to tackle the existing challenges and constraints is needed. Hence, India's approach to implementation of AI has to be guided by optimisation of social goods, rather than maximisation of topline growth.

From a technology perspective, the strategy is to maximise the late-movers' advantage. Acknowledging that India is some distance away from consistently delivering home grown pioneering technology solutions in AI, adapting and innovating the technology for India's unique needs and opportunities would help it in leap frogging, while simultaneously building the foundational R&D capability aimed at ensuring competitiveness in the long run.

Solving for India, given the complexity and multi-dimensional aspects of most of our economic and societal challenges, can easily be extended to the rest of the emerging and developing economies. An integral part of India's strategy for AI involves tackling common and complex global challenges that can be solved through technology intervention, and India's scale and opportunity landscape provides the ideal test-bed to ensure sustainable and scalable solutions.

The purpose of this paper is to lay the ground work for evolving the National Strategy for Artificial Intelligence. While this paper includes several recommendations, some of which may be deemed disruptive, specifics (e.g. execution and financial implications) have been consciously avoided, since wider consultations and consensus building is needed to refine these recommendations. This document is intended to serve as an "essential pre-read" in building a truly transformative approach in pursuit of #AIforAll.



Amitabh Kant

CEO, NITI Aayog

Executive Summary

India's Approach to Leadership in AI

AI refers to the ability of machines to perform cognitive tasks like thinking, perceiving, learning, problem solving and decision making. Initially conceived as a technology that could mimic human intelligence, AI has evolved in ways that far exceed its original conception. With incredible advances made in data collection, processing and computation power, intelligent systems can now be deployed to take over a variety of tasks, enable connectivity and enhance productivity. As AI's capabilities have dramatically expanded, so have its utility in a growing number of fields.

The truly transformative nature of the technology, yet the nascent stage of its adoption worldwide, provides India with an opportunity to define its own brand of AI leadership. #AIforAll - the brand proposed for India implies inclusive technology leadership, where the full potential of AI is realised in pursuance of the country's unique needs and aspirations. The strategy should strive to leverage AI for economic growth, social development and inclusive growth, and finally as a "Garage" for emerging and developing economies.

While AI has the potential to provide large incremental value to a wide range of sectors, adoption till date has been driven primarily from a commercial perspective. Technology disruptions like AI are once-in-a-generation phenomenon, and hence large-scale adoption strategies, especially national strategies, need to strike a balance between narrow definitions of financial impact and the greater good. NITI Aayog has decided to focus on five sectors that are envisioned to benefit the most from AI in solving societal needs:

a) Healthcare: increased access and affordability of quality healthcare,
b) Agriculture: enhanced farmers' income, increased farm productivity and reduction of wastage,
c) Education: improved access and quality of education,
d) Smart Cities and Infrastructure: efficient and connectivity for the burgeoning urban population, and
e) Smart Mobility and Transportation: smarter and safer modes of transportation and better traffic and congestion problems.

To truly reap the benefits of deploying AI at scale, the report identifies the following barriers that need to be addressed in order to achieve the goals of #AIforAll:

a) Lack of broad based expertise in research and application of AI,
b) Absence of enabling data ecosystems – access to intelligent data,
c) High resource cost and low awareness for adoption of AI,
d) Privacy and security, including a lack of formal regulations around anonymisation of data, and
e) Absence of collaborative approach to adoption and application of AI.

Superior research capabilities have been the cornerstone of leadership aspirations in emerging technologies and effectively realising the growth potential requires expertise in both core and applied research. Despite indications of recent positive efforts in this aspect of technology, AI research in India is still in its infancy and requires large scale concerted and collaborative interventions.

The paper proposes a two-tiered structure to address India's AI research aspirations:

a) *Centre of Research Excellence (CORE)* focused on developing better understanding of existing core research and pushing technology frontiers through creation of new knowledge;
b) *International Centers of Transformational AI (ICTAI)* with a mandate of developing and deploying application-based research. Private sector collaboration is envisioned to be a key aspect of ICTAIs.

The research capabilities are proposed to be complemented by an umbrella organisation responsible for providing direction to research efforts through analysis of socio-economic indicators, studying global advancements, and encouraging international collaboration. Pursuing "*moonshot research projects*" through specialised teams, development of a dedicated supranational agency to channel research in solving big, audacious problems of AI – "CERN for AI", and developing common computing and other related infrastructure for AI are other key components research suggested.

As technology increasingly disrupts the nature of jobs and shifts the benchmarks of technological aptitude, skilling and reskilling of workforce forms an integral part of our approach to adopting AI. There is an emergent need for reskilling the existing workforce and developing future talent in accordance with the changing needs of the job market. This could be done via the adoption of decentralised teaching mechanisms working in collaboration with the private sector and educational institutions to prescribe certification with value. Furthermore, promotion of job creation in new areas, like data annotation needs to be identified and promoted, as these would have the potential of absorbing a large portion of the workforce that may find itself redundant due to increasing automation.

Adoption of AI across the value chain viz. startups, private sector, PSUs and government entities, will truly unlock the potential by creating a virtuous cycle of supply and demand. The barriers to AI development and deployment can effectively be addressed by adopting the marketplace model – one that enables market discovery of not only the price but also of different approaches that are best suited to achieve the desired results. A three-pronged, formal marketplace could be created focusing on data collection and aggregation, data annotation and deployable models. There could be a common platform called the National AI Marketplace (NAIM).

Furthermore, for accelerated adoption of a highly collaborative technology like AI, the government has to play the critical role of a catalyst in supporting partnerships, providing access to infrastructure, fostering innovation through research and creating the demand by seeking solutions for addressing various governmental needs.

As AI-based solutions permeate the way we live and do business, questions on ethics, privacy and security will also emerge. Most discussions on ethical considerations of AI are a derivation of the FAT framework (Fairness, Accountability and Transparency). A consortium of Ethics Councils at each Centre of Research Excellence can be set up and it would be expected that all COREs adhere to standard practice while developing AI technology and products.

Data is one of the primary drivers of AI solutions, and thus appropriate handling of data, ensuring privacy and security is of prime importance. Challenges include data usage without consent, risk of identification of individuals through data, data selection bias and the resulting discrimination of AI models, and asymmetry in data aggregation. The paper suggests establishing data protection frameworks and sectorial regulatory frameworks, and promotion of adoption of international standards.

In order for India to ride the AI innovation wave, a robust intellectual property framework is required. Despite a number of government initiatives in strengthening the IP regime, challenges remain, especially in respect of applying stringent and narrowly focused patent laws to AI applications – given the unique nature of AI solution development. The importance of data to development of useful models is one such example. To tackle these issues, establishment of IP facilitation centers to help bridge the gap between practitioners and AI developers, and adequate training of IP granting authorities, judiciary and tribunals is suggested.

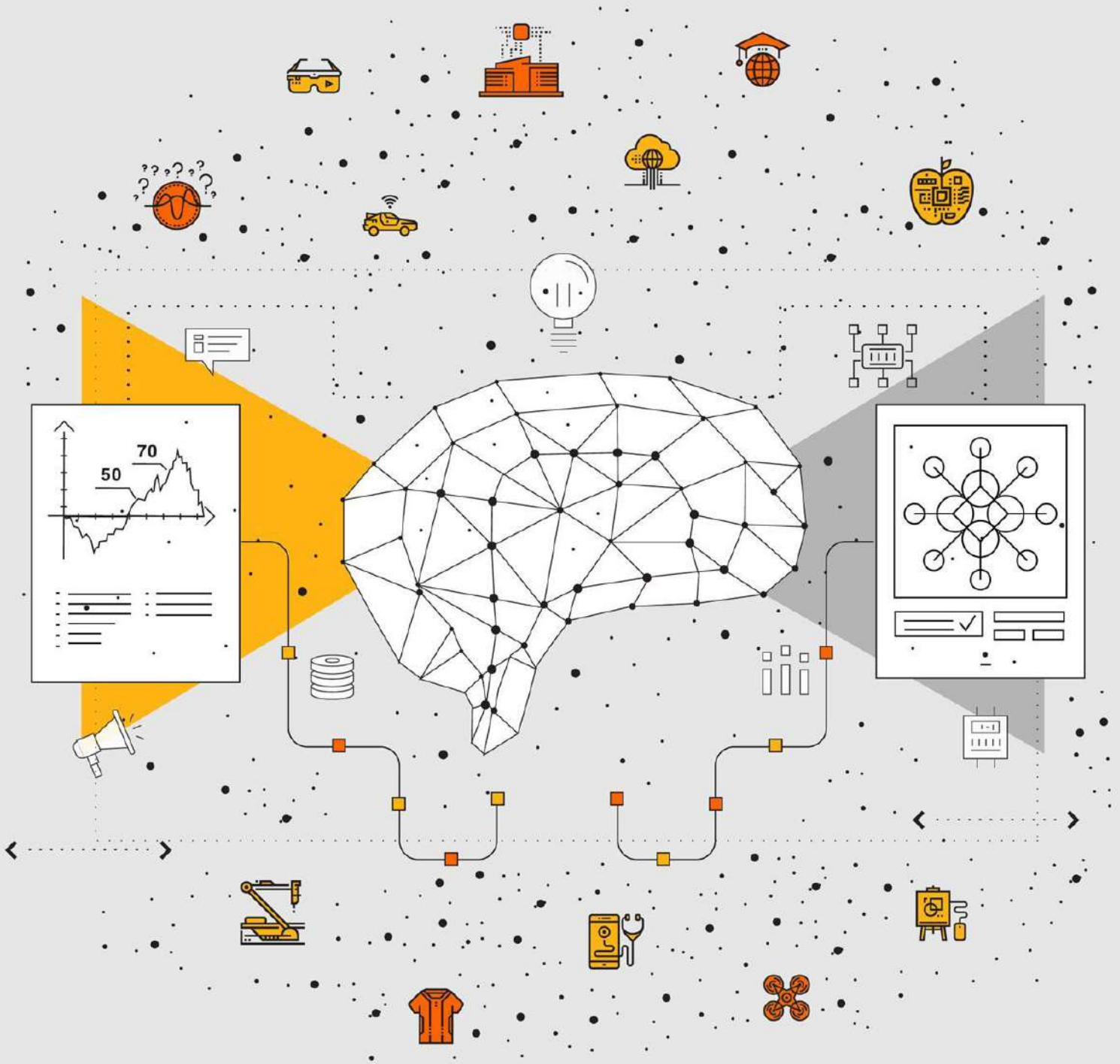
The AI strategy is aimed at primarily guiding an inevitable wave of change for quicker and better impact. The AI ecosystem is rapidly evolving and taking societies into uncharted territory. For now, we can begin to ask some of the big questions that each society must answer for itself: are we ready to manage data ethically? How do we bridge the digital divide? Which innovations are worthy of public funds and

partnerships? Bringing these questions into the open is the most important step in ensuring that AI advances create a better society.

There has been tremendous activity concerning AI policy in different countries over the past couple of years. Governments in USA, UK, France, Japan and China have released their policy and strategy papers relating to AI. In order to establish a leadership role, it is important for India to take the plunge and start by releasing a Strategy Paper to initiate the roll out of an ambitious programme that would ensure for India its rightful place in this transformational era.

This page has been intentionally left blank

PART I: SETTING THE CONTEXT



What is Artificial Intelligence?

A technical primer

AI might just be the single largest technology revolution of our live times, with the potential to disrupt almost all aspects of human existence. Andrew Ng, Co-founder of Coursera and formerly head of Baidu AI Group / Google Brain, compares the transformational impact of AI to that of electricity 100 years back. With many industries aggressively investing in cognitive and AI solutions, global investments are forecast to achieve a compound annual growth rate (CAGR) of 50.1% to reach USD57.6 billion in 2021[1].

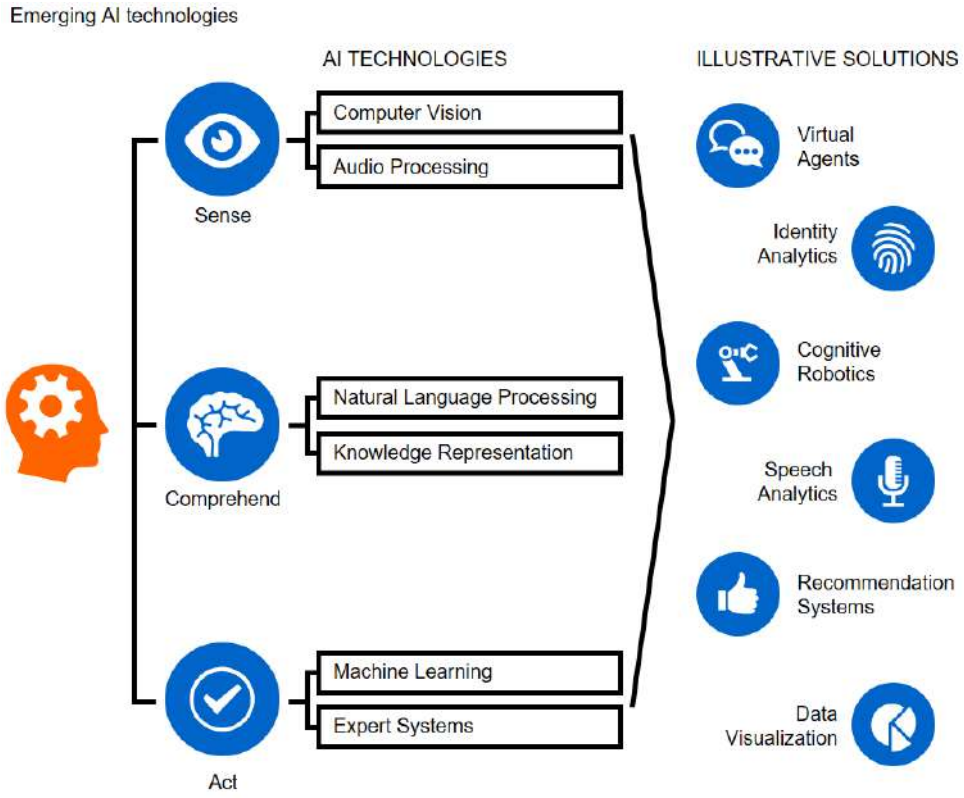
AI is not a new phenomenon, with much of its theoretical and technological underpinning developed over the past 70 years by computer scientists such as Alan Turing, Marvin Minsky and John McCarthy. AI has already existed to some degree in many industries and governments. Now, thanks to virtually unlimited computing power and the decreasing costs of data storage, we are on the cusp of the exponential age of AI as organisations learn to unlock the value trapped in vast volumes of data.

AI is a constellation of technologies that enable machines to act with higher levels of intelligence and emulate the human capabilities of sense, comprehend and act. Thus, computer vision and audio processing can actively perceive the world around them by acquiring and processing images, sound and speech. The natural language processing and inference engines can enable AI systems to analyse and understand the information collected. An AI system can also take action through technologies such as expert systems and inference engines or undertake actions in the physical world. These human capabilities are augmented by the ability to learn from experience and keep adapting over time. AI systems are finding ever-wider application to supplement these capabilities across enterprises as they grow in sophistication.

Irrespective of the type of AI being used, however, every application begins with large amounts of training data. In the past, this kind of performance was driven by rules-based data analytics programs, statistical regressions, and early “expert systems.” But the explosion of powerful deep neural networks now gives AI something a mere program doesn’t have: the ability to do the unexpected.

[1] *Worldwide Semi-annual Cognitive Artificial Intelligence Systems Spending Guide from International Data Corp. (IDC), 2017*

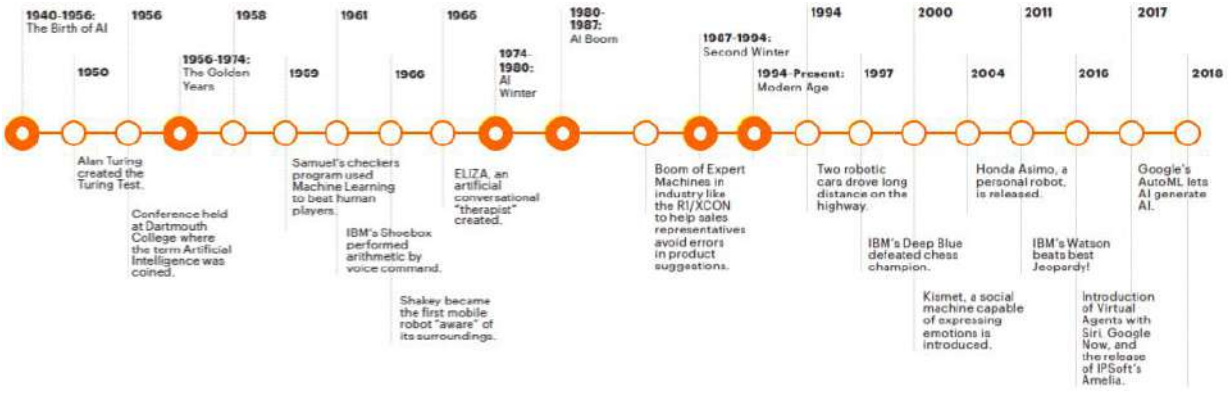
Figure 1: *What is Artificial Intelligence*



Source:
Accenture

AI technology has experienced a checkered history of waves of optimism followed by disappointment and periods of inertia, dubbed as "*AI winters*". Each previous breakthrough has only ever partly lived up to the hype it generated, and none has managed to kick-start the technology into the mainstream.

Figure 2: *Evolution of AI*



Source:
Accenture

The big change today is that we are in an unprecedented period of technology innovation across so many different fields that gives us the belief that the "*AI Spring*" has not only arrived but is here to stay. Key developments responsible for this optimism are:

a) *Unlimited access to computing power*: The worldwide public cloud services market is projected to grow 21.4% in 2018 to total USD186.4 billion, up from USD153.5 billion in 2017, according to Gartner, Inc. The access is amplified by rapid increase in computational power.
b) *Huge fall in cost of storing data*: We are in an age where the hard drive cost per gigabyte of data has been falling exponentially, to the extent that we are approaching near zero marginal cost for storing data (down from USD500,000 a gigabyte in 1980 to 2 cents a gigabyte in 2017).
c) *Explosion in data that is digitised*: As per IDC forecasts, by 2025, the global data sphere will grow to 163 zettabytes (that is a trillion gigabytes)[2], or ten times the 16.1ZB of data generated in 2016. As Barry Smyth, Professor of Computer science at University College Dublin, says: "Data is to AI what food is to humans." So, in a more digital world, the exponential growth of data is constantly feeding AI improvements.

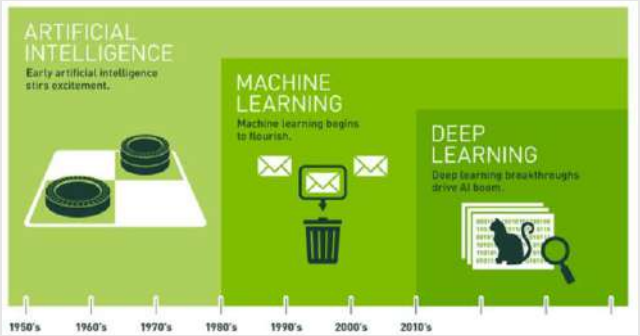
Consider, for example, the vastly increased processing power that comes from using Graphics Processing Units (GPUs) in place of Central Processing Units (CPUs). Google, in May 2017, announced that its Tensor Processing Unit (TPU) delivered 30-80 times higher performance-per-watt than contemporary CPUs and GPUs. When you add the decreasing cost of storage to the mix, plus the exponential growth in data volumes, together with the emergence of open source platforms and frameworks, you have got a uniquely potent combination of technologies and capabilities. It all adds up to a very powerful foundation to give AI its critical mass for mainstream adoption.

Box 1: Machine Learning and Deep Learning

Machine Learning, a term coined by Artur Samuel in 1959, meant "the ability to learn without being explicitly programmed." Machine Learning involves the use of algorithms to parse data and learn from it, and making a determination or prediction as a result. Instead of hand coding software libraries with well-defined specific instructions for a particular task, the machine gets "trained" using large amounts of data and algorithms, and in turn gains the capability to perform specific tasks.

"Deep Learning is a technique for implementing Machine Learning. Deep Learning was inspired by the structure and function of the brain, specifically the interconnecting of many neurons. Artificial Neural Networks (ANNs) are algorithms that are based on the biological structure of the brain. In ANNs, there are 'neurons' which have discrete layers and connections to other "neurons". Each layer picks out a specific feature to learn. It's this layering that gives deep learning its name, depth is created by using multiple layers as opposed to a single layer[3]."

Figure 3: *AI, ML and Deep Learning*



Source: nVIDIA

[2] *Data Age 2025: "The Evolution of Data to Life-Critical whitepaper by International Data Corporation, 2017"*

[3] *Medium.com: "The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning"*

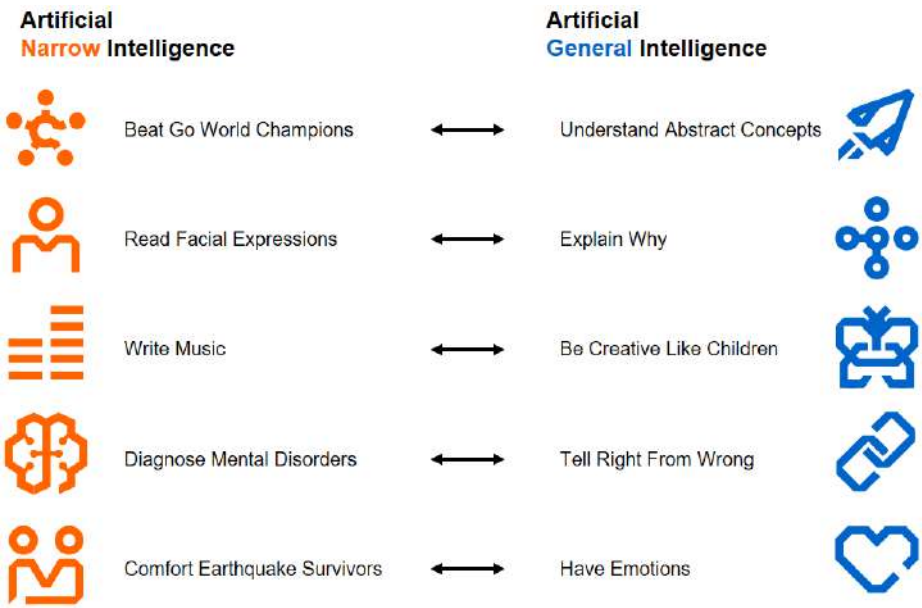
For a detailed analysis of major types of Machine Learning, algorithms and use cases, please refer to *Appendix I: Artificial Intelligence Explained.*

AI gets categorised in different ways and it may be useful to understand the various categories, their rationale and the implications.

a) *Weak AI vs. Strong AI*: Weak AI describes "simulated" thinking. That is, a system which appears to behave intelligently, but doesn't have any kind of consciousness about what it's doing. For example, a chatbot might appear to hold a natural conversation, but it has no sense of who it is or why it's talking to you. Strong AI describes "actual" thinking. That is, behaving intelligently, thinking as human does, with a conscious, subjective mind. For example, when two humans converse, they most likely know exactly who they are, what they're doing, and why.
b) *Narrow AI vs. General AI*: Narrow AI describes an AI that is limited to a single task or a set number of tasks. For example, the capabilities of IBM's Deep Blue, the chess playing computer that beat world champion Gary Kasparov in 1997, were limited to playing chess. It wouldn't have been able to win a game of tic-tac-toe - or even know how to play. General AI describes an AI which can be used to complete a wide range of tasks in a wide range of environments. As such, it's much closer to human intelligence.
c) Superintelligence: The term "superintelligence" is often used to refer to general and strong AI at the point at which it surpasses human intelligence, if it ever does.

While big strides have been made in Artificial Narrow Intelligence – algorithms that can process documents, drive vehicles or beat champion chess players, no one has yet claimed the first production or development of General AI. The weight of expert opinion is that we are a long way off the emergence of General AI.

Figure 4: *Narrow AI vs. General AI*



Source: Accenture

Global Developments in Artificial Intelligence

Benchmarking select countries

Countries around the world are becoming increasingly aware of the potential economic and social benefits of developing and applying AI. For example, China and U.K. estimate that 26% and 10% of their GDPs respectively in 2030 will be sourced from AI-related activities and businesses. There has been tremendous activity concerning AI policy positions and the development of an AI ecosystem in different countries over the last 18 to 24 months – the US published its AI report in December 2016; France published the AI strategy in January 2017 followed by a detailed policy document in March 2018; Japan released a document in March 2017; China published the AI strategy in July 2017; and U.K. released its industrial strategy in November 2017.

Infrastructural supply side interventions have been planned by various countries for creating a larger ecosystem of AI development. Creation of “data trusts”, rolling out of digital connectivity infrastructure such as 5G / full fiber networks, common supercomputing facilities, fiscal incentives and creation of open source software libraries are some of the focus areas of various governments as committed in their strategy papers.

In the area of core research in AI and related technologies, universities and research institutions from the US, China and Japan have led the publication volume on AI research topics between 2010 and 2016. Universities in USA, primarily Carnegie Mellon University, Massachusetts Institute of Technology and Stanford, took an early lead in AI research by offering new courses, establishing research facilities and instituting industry partnerships. Off late, Chinese universities, especially Peking and Tsinghua Universities have caught on to the race by utilising large scale public funding and extensive research partnerships with private companies.

For building the future workforce for AI, countries are also significantly increasing the allocation of resources for Science, Technology, Engineering and Maths (STEM) talent development through investment in universities, mandating new courses (e.g., AI and law), and offering schemes to retrain people. For instance, U.K. has planned to build over 1,000 government supported PhD researchers by 2025 and set up a Turing fellowship to support an initial cohort of AI fellows while China has launched a five-year university program to train at least 500 teachers and 5,000 students working on AI technologies.

Governance structures for enabling all the above mandates vary across countries. Many countries have instituted dedicated public offices such as Ministry of AI (UAE), and Office of AI and AI Council (U.K.) while China and Japan have allowed existing ministries to take up AI implementation in their sectoral areas. Not just national governments, but even local city governments have become increasingly aware about the importance and potential of AI and have committed public investments.

National governments have significantly increased public funding for AI through commitments such as increasing the R&D spend, setting up industrial and investment funds in AI startups, investing in network and infrastructure and AI-related public procurements. China, USA, France and Japan have committed significant public spending for AI technology development and adoption.

These countries are also leveraging different combinations of public-private-academia to develop and promote AI. Development of technology parks, and connecting large corporations with startups and

forming “national teams” with large private players to undertake fundamental and applied research are some of the public-private partnership approaches various national governments have espoused.

AI technology development and applications are evolving rapidly with major implications for economies and societies. A study by EY and NASSCOM found that by 2022, around 46% of the workforce will be engaged in entirely new jobs that do not exist today, or will be deployed in jobs that have radically changed skillsets[4]. If some countries decide to wait for a few years to establish an AI strategy and put in place the foundations for developing the AI ecosystem, it seems unlikely that they would be able to attain and match up to the current momentum in the rapidly changing socio-economic environment. Therefore, the need of the hour is to develop a policy framework that will help set up a vibrant AI ecosystem in India.

A detailed study of various country strategies for AI is placed in the *Appendix II: Global Country Strategy Review*.

[4] *Future of Jobs in India: A 2022 Perspective, 2017*

Artificial Intelligence and India

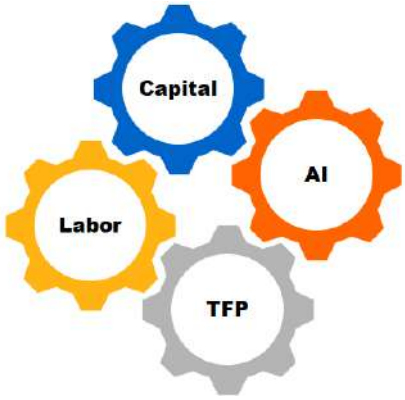
Identifying priority areas for India's efforts in Artificial Intelligence

A national AI strategy needs to be premised on a framework which is adapted to India's unique needs and aspirations, while at the same time, is capable of achieving the country's full potential of leveraging AI developments. Such a framework could be seen as an aggregation of the following three distinct, yet inter-related components:

a) *Opportunity*: the economic impact of AI for India
b) *AI for Greater Good*: social development and inclusive growth
c) *AI Garage for 40% of the world*: solution provider of choice for the emerging and developing economies (ex-China) across the globe

Opportunity: the economic impact of Artificial Intelligence for India

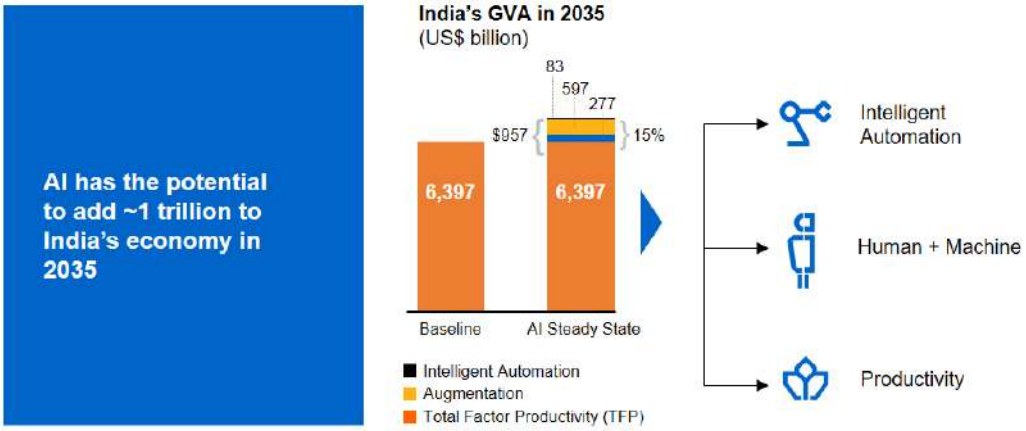
AI is emerging as a new factor of production, augmenting the traditional factors of production viz. labor, capital and innovation and technological changes captured in total factor productivity. AI has the potential to overcome the physical limitations of capital and labour, and open up new sources of value and growth. From an economic impact perspective, AI has the potential to drive growth through enabling: (a) intelligent automation i.e. ability to automate complex physical world tasks that require adaptability and agility across industries, (b) labour and capital augmentation: enabling humans to focus on parts of their role that add the most value, complementing human capabilities and improving capital efficiency, and (c) innovation diffusion i.e. propelling innovations as it diffuses through the economy. AI innovations in one sector will have positive consequences in another, as industry sectors are interdependent based on value chain. Economic value is expected to be created from the new goods, services and innovations that AI will enable.



Accenture, in its recent AI research reports[5], provides a framework for evaluating the economic impact of AI for select G20 countries and estimates AI to boost India's annual growth rate by 1.3 percentage points by 2035.

[5] *Rewire for Growth: Accelerating India's Economic Growth with Artificial Intelligence, Accenture*

Figure 5: *Unlocking innovation through AI*



Source: Accenture

AI for Greater Good: social development and inclusive growth

Beyond just the headline numbers of economic impact, a disruptive technology such as AI needs to be seen from the perspective of the transformative impact it could have on the greater good – improving the quality of life and access of choice to a large section of the country. In that sense, the recent advancements in AI seem to be custom-made for the unique opportunities and challenges that India faces. Increased access to quality health facilities (including addressing the locational access barriers), inclusive financial growth for large sections of population that have hitherto been excluded from formal financial products, providing real-time advisory to farmers and help address unforeseen factors towards increasing productivity, building smart and efficient cities and infrastructure to meet the demands of rapidly urbanising population are some of the examples that can be most effectively solved through the non-incremental advantages that a technology such as AI can provide.

AI Garage for 40% of the world

In addition to providing unique opportunities, India provides a perfect "*playground*" for enterprises and institutions globally to develop scalable solutions which can be easily implemented in the rest of the developing and emerging economies. Simply put, *Solve for India* means solve for 40% or more of the world. An advanced AI based solution for early diagnosis of tuberculosis (one of the top-10 causes of deaths worldwide), for example, could easily be rolled out to countries in South East Asia or Africa, once developed and refined in India. Beyond healthcare, AI technologies in the other sectors including agriculture, education and mobility are set to transform the world. The commonality of issues with regard to the above sectors across developing countries provides the ideal use case of developing AI solutions that could be adapted for multiple markets. Hence, AI technologies suited for the Indian agricultural sector could easily be customised for other developing nations based on their local climatic conditions. Education continues to be a major concern in almost all developing countries. AI technologies that are capable of imparting quality education to India's linguistically diverse population could prove very useful in other developing nations.

Another aspect of India's potential as a leader in AI is it proven track record in technology solution provider of choice. *Solved in India* (or more accurately, solved by Indian IT companies) could be the model going forward for Artificial Intelligence as a Service (AIaaS). Indian IT companies have been pioneers in bringing technology products and developments as solutions across the globe. As AI matures and generalised applications become common place, its advantage India when it comes to large scale implementation. Furthermore, India's competence in IT combined with opportunities, such as interoperability between

multiple languages, provides the much needed impetus for finding scalable solutions for problems that have global implications, such as NLP.

Artificial Intelligence has the potential to provide large incremental value to a wide range of sectors globally, and is expected to be the key source of competitive advantage for firms.

a) *Healthcare*: Application of AI in healthcare can help address issues of high barriers to access to healthcare facilities, particularly in rural areas that suffer from poor connectivity and limited supply of healthcare professionals. This can be achieved through implementation of use cases such as AI driven diagnostics, personalised treatment, early identification of potential pandemics, and imaging diagnostics, among others.

b) *Agriculture*: AI holds the promise of driving a food revolution and meeting the increased demand for food (global need to produce 50% more food and cater to an additional 2 billion people by 2050 as compared to today). It also has the potential to address challenges such as inadequate demand prediction, lack of assured irrigation, and overuse / misuse of pesticides and fertilisers. Some use cases include improvement in crop yield through real time advisory, advanced detection of pest attacks, and prediction of crop prices to inform sowing practices.

c) *Smart Mobility, including Transports and Logistics*: Potential use cases in this domain include autonomous fleets for ride sharing, semi-autonomous features such as driver assist, and predictive engine monitoring and maintenance. Other areas that AI can impact include autonomous trucking and delivery, and improved traffic management.

d) *Retail*: The retail sector has been one of the early adopters of AI solutions, with applications such as improving user experience by providing personalised suggestions, preference-based browsing and image-based product search. Other use cases include customer demand anticipation, improved inventory management, and efficient delivery management.

e) *Manufacturing*: Manufacturing industry is expected to be one of the biggest beneficiaries of AI based solutions, thus enabling 'Factory of the Future' through flexible and adaptable technical systems to automate processes and machinery to respond to unfamiliar or unexpected situations by making smart decisions. Impact areas include engineering (AI for R&D efforts), supply chain management (demand forecasting), production (AI can achieve cost reduction and increase efficiency), maintenance (predictive maintenance and increased asset utilisation), quality assurance (e.g. vision systems with machine learning algorithms to identify defects and deviations in product features), and in-plant logistics and warehousing.

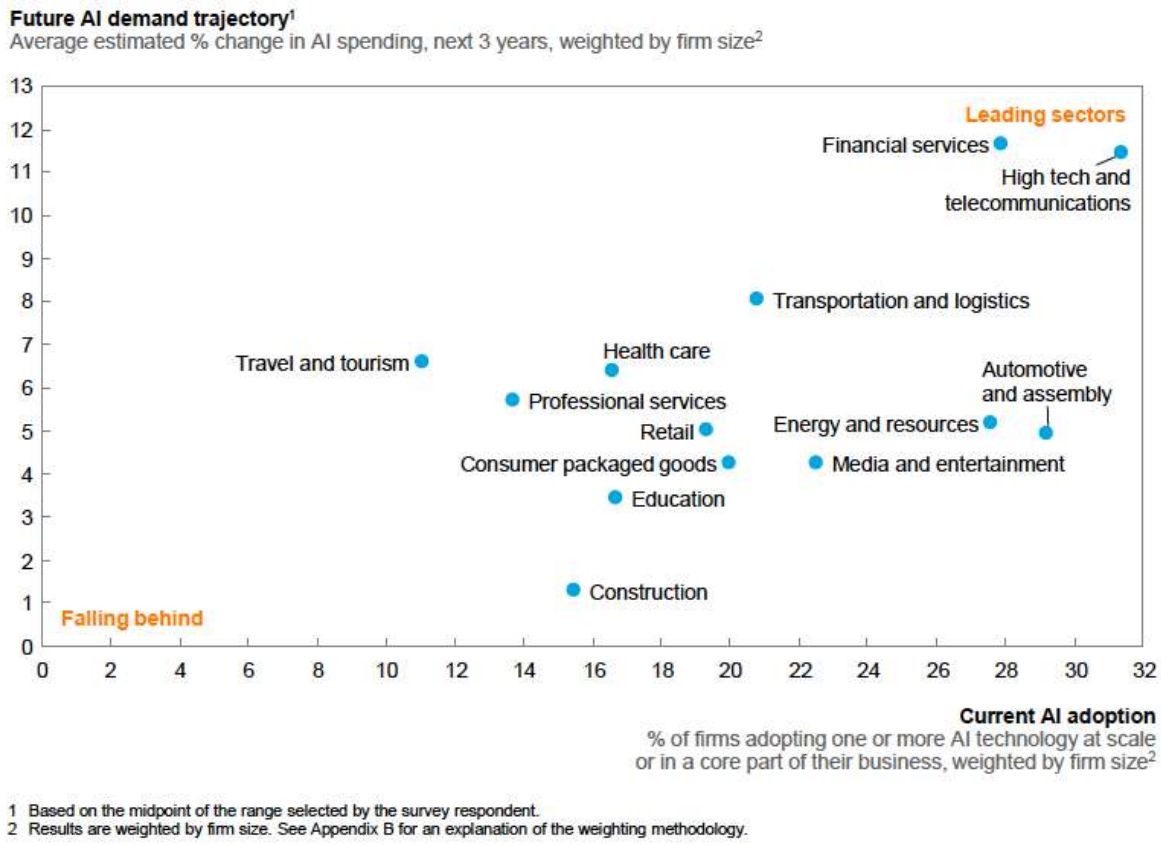
f) *Energy*: Potential use cases in the energy sector include energy system modelling and forecasting to decrease unpredictability and increase efficiency in power balancing and usage. In renewable energy systems, AI can enable storage of energy through intelligent grids enabled by smart meters, and also improve the reliability and affordability of photovoltaic energy. Similar to the manufacturing sector, AI may also be deployed for predictive maintenance of grid infrastructure.

g) *Smart Cities*: Integration of AI in newly developed smart cities and infrastructure could also help meet the demands of a rapidly urbanising population and providing them with enhanced quality of life. Potential use cases include traffic control to reduce congestion and enhanced security through improved crowd management.

h) *Education and Skilling*: AI can potentially solve for quality and access issues observed in the Indian education sector. Potential use cases include augmenting and enhancing the learning experience through personalised learning, automating and expediting administrative tasks, and predicting the need for student intervention to reduce dropouts or recommend vocational training.

Adoption of AI by various sectors have been influenced by, among other factors, technical and regulatory challenges, but commercial implications has been the biggest determinant. While technical feasibility, availability of structured data, regulatory barriers, privacy considerations, ethical issues, preference for human relationship have all played their roles in determining the readiness of a sector for large scale AI adoption; compelling business use cases (e.g. improved efficiency, accuracy, speed, forecasting and accurate decision making) that lead to direct impact on revenue and profitability have been the biggest driver for companies to pursue accelerated adoption of AI. As illustrated in McKinsey Global Institute's AI adoption and use survey, sectors leading the AI adoption today also intend to grow their investment in AI the most, thus further reinforcing the varying degrees of AI adoption across sectors.

Figure 6: *Current AI adoption and future AI investments by sector*



1 Based on the midpoint of the range selected by the survey respondent.
2 Results are weighted by firm size. See Appendix B for an explanation of the weighting methodology.

Source: McKinsey Global Institute AI adoption and use survey

It comes as no surprise that Banking and Financial Services sector has been one of the leading sectors globally when it comes to AI adoption, and India has also seen a steep increase in AI based implementation in recent times. Existing and potential use of Artificial Intelligence in this sector include improved customer interaction through personalised engagement, virtual customer assistance, and chatbots; improved processes through deployment of intelligent automation in rule based back-office operations; development of credit scores through analysis of bank history or social media data; and fraud analytics for proactive monitoring and prevention of various instances of fraud, money laundering, malpractice, and the prediction of potential risks. AI in this sector has also been employed in wealth management viz. robo-advisory, algorithmic trading and automated transactions.

Similarly, manufacturing sector, primarily automotive and assembly, has been one of the first sectors to implement advanced robotics at scale. The manufacturing sector in India hasn't been far behind, as reflected in a recent study by BCG, where India was ranked 3rd in the world in AI implementation in

manufacturing, ahead of nations such as Germany, with 19% of companies in the sector already using AI to a significant extent[6].

These trends have also been reflected in the nature of investment in research in India, with private sector initiatives such as the Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI), choosing to focus their efforts in applied research on sectors such as manufacturing analytics and financial analytics.

Figure 6 also reveals that sectors like Healthcare and Education have quite a lot of ground to cover as far as AI adoption is concerned. Healthcare, despite being one of the hottest areas of AI startup investments (*Appendix IV: What Do the Markets Say*), is tricky, especially in the Indian context. Agriculture doesn't even feature in the analysis above. Another analysis by McKinsey Global Institute indicates that potential value of AI for agriculture was in the bottom tercile of 19 sectors evaluated[7], and could be a possible explanation for diminished private sector led AI adoption in agriculture. In sectors such as these, externalities from adoption of AI far outweigh the economic returns realised by private initiatives, and hence the role of government becomes pivotal in ensuring large scale AI intervention.

NITI Aayog has evaluated various sectors that will be impacted by AI and has taken a conscious decision to focus on a select set of sectors where only private sector led initiatives will not lead to achieving desired societal outcomes. In addition to Healthcare and Agriculture, focus sectors include Education (preparing tomorrow's generation to leverage the global AI revolution to India's advantage), Smart Cities and Infrastructure (solving for India's rapidly urbanising population) and Smart Mobility and Transportation (solving for challenges congestion, pollution, high rates of road accidents leading to economic inefficiency and enormous human cost).

An unrelated but interesting paradigm for AI application is the "AI + X" approach. Despite its vast potential, the capabilities of AI today are limited to tasks for which it has been specifically trained, and are still many years from achieving human like consciousness. AI today should thus be regarded as an enhancement, or enabler of increased efficiency in previously existing processes, rather than capable of a complete overhaul of traditional tasks. Deployment can be viewed through the paradigm of "take an existing process, and add AI" or "AI + X"; where "X" can range from tasks such as driving a car, where AI can provide incremental value through improved routing and energy management, to act of sowing seeds, where AI can help inform decision making and improve productivity.

Similar to the effects of electricity, AI can increasingly be seen as an intelligent, additive utility that can be deployed at will, but remain largely invisible to the tasks performer. This vision is perhaps best put by author Kevin Kelley:

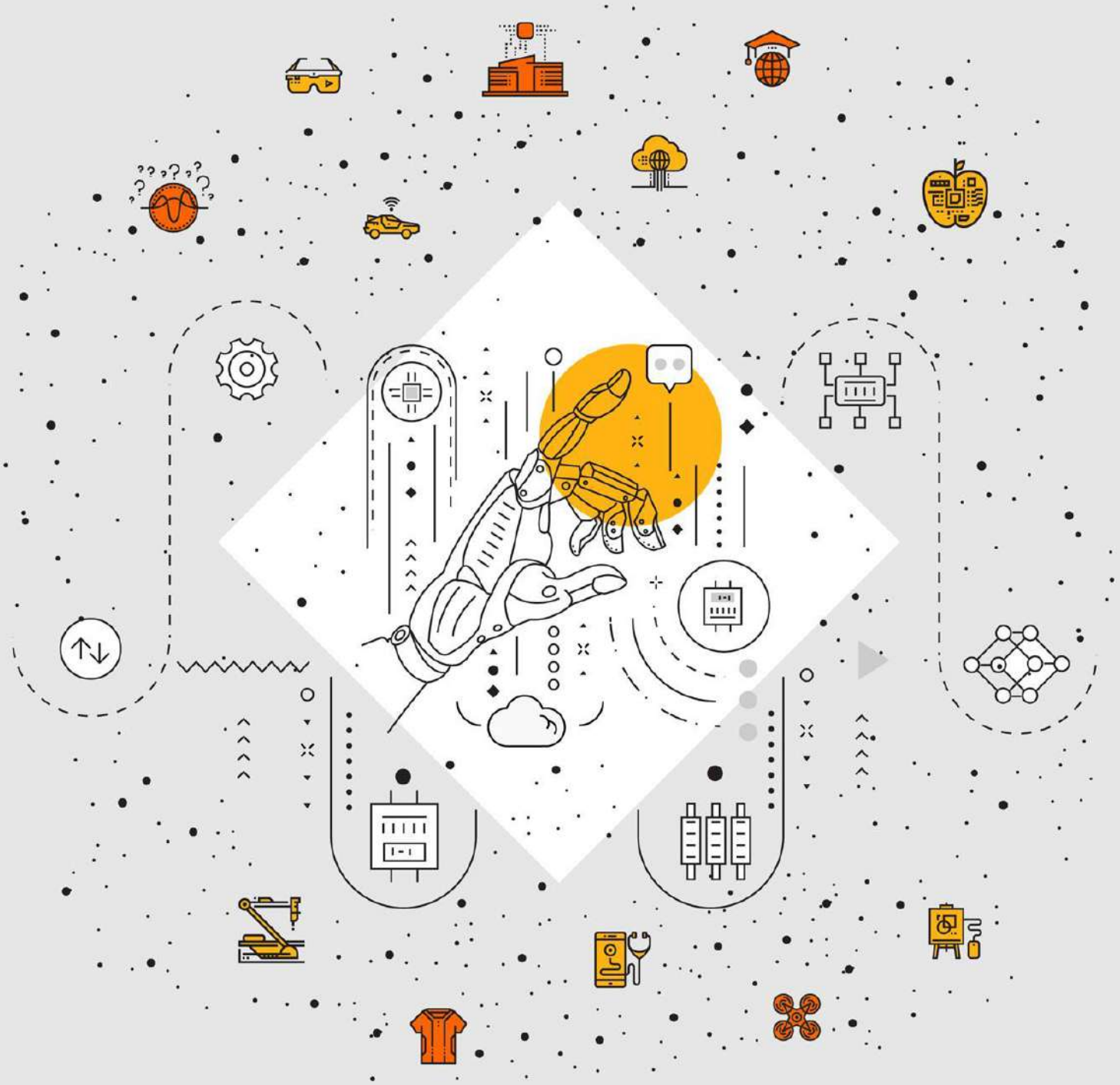
"*There is almost nothing we can think of that cannot be made new, different, or more valuable by infusing it with some extra IQ. In fact, the business plans for the next 10,000 startups are easy to forecast: Take X and add AI*" .

In applying this paradigm to the development of a national strategy, it is thus important to consider the challenges faced by individual sectors, or various manifestations of "X" to best identify the incremental value that AI can provide. The paradigm also cements the need for collaboration with sectorial stakeholders in the application of the technology. The paradigm provides a useful framework to analyse what is possible in terms of technology intervention today.

[6] *BCG "AI in the Factory of the Future"*

[7] *McKinsey Global Institute "Notes From The AI Frontier: Insights From Hundreds Of Use Cases"*

PART II:
APPLICATION AREAS AND ENABLERS



Focus areas for AI intervention

Sectoral deep dives

Healthcare

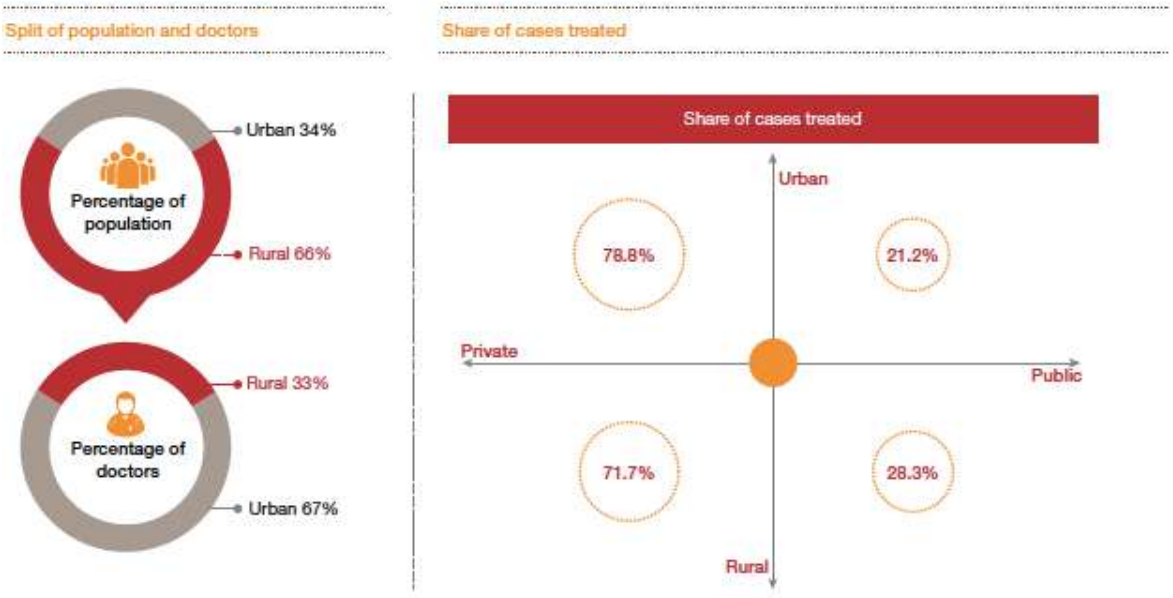
Healthcare is one of the most dynamic, yet challenging, sectors in India, and is expected to grow to USD280 billion by 2020, at a CAGR of upwards of 16%, from the current ~USD100 billion[8].

Yet, it faces major challenges of quality, accessibility and affordability for a large section of the population:

a) *Shortage of qualified healthcare professionals and services like qualified doctors, nurses, technicians and infrastructure*: as evidenced in 0.76 doctors and 2.09 nurses per 1,000 population (as compared to WHO recommendations of 1 doctor and 2.5 nurses per 1,000 population respectively) and 1.3 hospital beds per 1,000 population as compared to WHO recommended 3.5 hospital beds per 1,000 population[9].

b) N*on-uniform accessibility to healthcare across the country with physical access* continuing to be the major barrier to both preventive and curative health services, and glaring disparity between rural and urban India.

Figure 7: *Accessibility of Healthcare across India*



Source: PwC Analysis, World Bank data (2017)

[8] *FICCI-KPMG study*

[9] *WHO website, PwC analysis*

With most of the private facilities concentrated in and around tier 1 and tier 2 cities, patients have to travel substantial distances for basic and advanced healthcare services. (Box 2: What TMH's Cancer heat map tells us about the availability of healthcare in India?)

The problem is further accentuated by lack of consistent quality in healthcare across India, most of the services provided is individual driven rather than institution driven, and less than 2% of hospitals in India are accredited.

Box 2: What TMH's Cancer heat map tells us about the availability of healthcare in India?

Tata Memorial Hospital, one of the leading cancer hospitals in India, registered more than 67,000 new registrations for cancer treatment in 2015. While the hospital is located in Mumbai, less than 23% of the new patients were geographically based in Maharashtra, with a whopping 21.7% of patients traveling from the states of UP, Bihar, Jharkhand and West Bengal to TMH.

***Figure 8**: Geographic location of TMH cancer patients*

That these patients had to travel more than 1,800 km, on an average, to avail cancer treatment is an unfortunate tale of lack of access to quality healthcare. In addition to battling a potentially life threatening disease, the patients are saddled by the stress and financial implications of traveling long way away from home. While the data is not available to such an effect, it wouldn't be surprising to find that most of these patients choose to travel to TMH when cancer has developed to an advanced stage, thus further reducing the chances of successful cure and treatment.

***Credit**: Tata Memorial Centre*

c) *Affordability* remains a problem with private expenditure accounting for ~70% of healthcare expenses, of which ~62% is out-of-pocket expenditure, probably one of the highest in the world. Significant portion of hospital costs in both rural (~47%) and urban India (~31%) are financed by loans and sale of assets. Poor and marginalised are hit the most, and as per the Government estimates, a sizeable part of the population (~63 million) are faced with poverty every year because of their healthcare expenditure[10].

d) *Reactive approach to essential healthcare* largely due to lack of awareness, access to services and behavioral factors implies that majority of patients approach a hospital / physician only when a disease has reached an advanced stage, thus increasing the cost of care and reducing the chances of recovery.

The Government of India has been making a series of large scale interventions to address India's healthcare challenges, viz. transformation of 1.5 lakh Health and Wellness Centers, developing district hospitals to cater to long-term care for non-communicable diseases, Ayushman Bharat Mission, promoting e-Health etc.

Box 3: Government of India's push for Universal Healthcare Coverage

The Government of India, through its recent policy interventions, has shown a bold commitment to achieve Universal Health Coverage and increased access to comprehensive primary health care. Through the Ayushman Bharat programme announced in Union Budget 2018, probably the world's largest government funded health care programme, the Government of India has embarked on a path breaking journey to ensure the affordability and accessibility of healthcare in India. The Ayushman Bharat – National Health Protection Mission (AB – NHPM) aims to provide insurance cover of INR 5 lakh per family per year for secondary and tertiary care hospitalisation. Ayushman Bharat is targeted at more than 10 crore families (approximately 50 crore beneficiaries / ~40% of India's population) belonging to the poor and vulnerable sections based on the SECC database, and doesn't impose any limitations on family size or age limit for the beneficiaries to avail benefits. The benefits package covers most medical and surgical conditions with minimal exclusions, covers pre and post hospitalisation expenses, and covers all pre-existing conditions from day one – thus simplifying availing requisite healthcare by the beneficiaries. The benefits of the Mission will be available at public hospitals as well as empaneled private health care facilities.

The Union Budget 2018 also included a commitment of ~INR1,200 crore for Health and Wellness Centres (HWC), which will lay the foundation for India's health system as envisioned in the National Health Policy 2017. These HWCs, to be set up by transforming 1.5 lakh Health Sub Centres from 2018 to 2022, are aimed at shifting primary healthcare from selective (reproductive and child health / few infectious diseases) to comprehensive (including screening and management of NCDs; screening and basic management of mental health ailments; care for common ophthalmic and ENT problems; basic dental health care; geriatric and palliative health care, and trauma care and emergency care). NCDs account for ~60% of mortality in India, 55% of which is premature. NCDs are predominantly chronic conditions and impact the poor most adversely, given the high costs of treatment involved. Prevention and early detection are therefore of the essence in reducing the disease burden attributable to these conditions as well as ensuring long-term follow-up and management of symptoms for patients. The HWCs, under the new implementation plan, will provide 12 basic healthcare services, expanding from the current package of 6 services. Crucially, these centres will provide preventive services to improve healthy behaviours for

[10] *National Health Policy 2015 Draft*

family health and control the incidence of communicable and non-communicable diseases among the population covered by HWCs. A key component of HWCs will be universal screening for NCDs. Screening for five NCDs and associated risk factors has been prioritised given the high burden of disease associated with them. These include hypertension, diabetes, as well as three common cancers - oral, breast and cervical. Screening for other conditions such as Chronic Obstructive Disease will be added subsequently. The HWCs will be operated by a mid-level health service provider, auxiliary nurse midwives, accredited social health activists and a male health worker responsible for comprehensive primary health care services for a population of about 5,000.

***Figure 9**: Features of HWC*

The NHPM and HWC, in unison, are aimed at holistically addressing the health needs of the population, including health promotion and disease prevention as well as the delivery of primary, secondary and tertiary services.

In addition, the government aims at leveraging technology to improve healthcare facilities through the:

a) the *National eHealth Authority (NeHA)* which will strategise eHealth adoption, define standards and a framework for the health sector, put in place electronic health exchanges for interoperability,

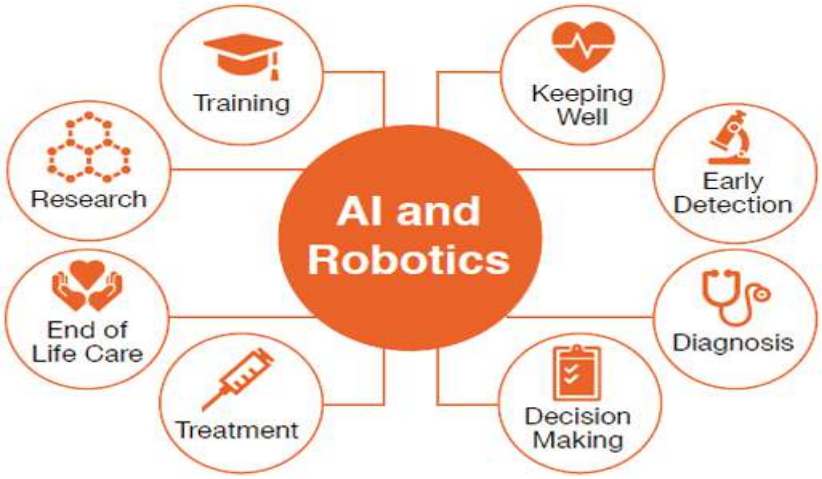
b) the *Integrated Health Information Program* (IHIP) to provide EHR to all citizens of India and provide interoperability to existing EHR/EMRs,

c) the *Electronic Health Record Standards* for India

Despite the obvious economic potential, the healthcare sector in India remains multi-layered and complex, and is ripe for disruption from emerging technologies at multiple levels. It is probably the most intuitive and obvious use case primed for intervention by AI driven solutions, as evidenced by the increasing activity from large corporates and startups alike in developing AI focused healthcare solutions. Adoption of AI for healthcare applications is expected to see an exponential increase in next few years. The healthcare market globally driven by AI is expected to register an explosive CAGR of 40% through

2021, and what was a USD600 million market in 2014 is expected to reach USD6.6 billion by 2021[11]. The increased advances in technology, and interest and activity from innovators, provides opportunity for India to solve some of its long existing challenges in providing appropriate healthcare to a large section of its population. AI combined with robotics and Internet of Medical Things (IoMT) could potentially be the new nervous system for healthcare, presenting solutions to address healthcare problems and helping the government in meeting the above objectives.

Figure 10: Potential use cases of AI in Healthcare



Source: PWC, "No longer science fiction, AI and robotics are transforming healthcare"

AI solutions can augment the scarce personnel and lab facilities; help overcome the barriers to access and solve the accessibility problem; through early detection, diagnostic, decision making and treatment, cater to a large part of India.

Cancer screening and treatment is an area where AI provides tremendous scope for targeted large scale interventions. India sees an incidence of more than 1 million new cases of cancer every year, and early detection and management can be crucial in an optimum cancer treatment regimen across the country. NITI Aayog is in an advanced stage for launching a programme to develop a national repository of annotated and curated pathology images. Another related project under discussions is an Imaging Biobank for Cancer.

Box 4: AI for India's cancer woes

Cancer screening and treatment is an area where AI provides tremendous scope for targeted large scale interventions. India sees an incidence of more than 1 million new cases of cancer every year, a number that is likely to increase given the increasing age of Indian population and lifestyle changes. Early detection and management can be crucial in an optimum cancer treatment regimen across the country. Good quality pathology service is the essential building block of cancer care, which unfortunately is not easily available outside select Indian cities. For an annual incidence of more than 1 million new cancer diagnosis every year, India has barely 2,000 pathologists experienced in oncology, and less than 500

[11] Frost & Sullivan, "From $600 M to $6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare"

pathologists who could be considered an expert oncopathologist. Machine learning solutions aimed at assisting a general pathologist in making quality diagnosis can very well plug this gap in providing essential healthcare. An essential pre-requisite in implementation of such a solution is availability of quality annotated pathology datasets. NITI Aayog is in an advanced stage for launching a programme to develop a national repository of annotated and curated pathology images. The components of such a repository include a move towards “Digital Pathology”, which entails all glass slides generated being scanned at high resolution and magnification, followed by accurate, precise and comprehensive annotation of the scanned images using various data sources & levels of clinical & pathological (gross pathology, histopathology and molecular) information available from day-to-day patient care.

Another related project under discussions is an Imaging Biobank for Cancer. Human cancers exhibit strong phenotypic differences that may be visualised noninvasively by expert radiologists (using imaging modalities). Recent literature suggests that certain image based features may correlate to molecular and clinical features like known mutations (KRAS, EGFR, etc.), receptor status, prognostic power, intra-tumor heterogeneity, gene expression patterns, etc. Reports have shown an association between radiographic imaging phenotypes and tumor stage, metabolism, hypoxia, angiogenesis and the underlying gene and/or protein expression profiles. These correlations, if rigorously established, may have a huge clinical impact as imaging is routinely used in clinical practice. Moreover, this provides an unprecedented opportunity to use artificial intelligence to improve decision-support in cancer treatment at low cost especially in countries like India. AI based Radiomics is an emerging field that refers to the comprehensive quantification of tumor phenotypes by applying a large number of quantitative imaging features. It has resulted in improvement to existing biomarker signature panels by adding imaging features.

***Credit**: Tata Memorial Centre for developing the concepts for Digital Pathology and Imaging Biobank*

NITI Aayog is working with Microsoft and Forus Health to roll out a technology for early detection of diabetic retinopathy as a pilot project. 3Nethra, developed by Forus Health, is a portable device that can screen for common eye problem. Integrating AI capabilities to this device using Microsoft’s retinal imaging APIs enables operators of 3Nethra device to get AI-powered insights even when they are working at eye checkup camps in remote areas with nil or intermittent connectivity to the cloud. The resultant technology solution also solves for quality issues with image capture and systems checks in place to evaluate the usability of the image captured.

AI based healthcare solutions can also help in making healthcare services more proactive – moving from “sick” care to true “health” care, with emphasis on preventive techniques.

Agriculture

While India has come a long way from being categorised as purely an agrarian economy, agriculture and allied sector still accounts for 49% of India's workforce, 16% of the country's gross domestic product (GDP)[12], and ensures food security to roughly 1.3 billion people.

Agriculture and allied sector is critical to India's growth story. To achieve and maintain an annual growth rate of 8 –10% for the Indian economy, agriculture sector must grow 4% or higher rate. The Government of India has recently prioritised *Doubling Farmers' Income as a National Agenda*; putting considerable focus on supply chain perspectives in agriculture and market development in addition to productivity augmentation.

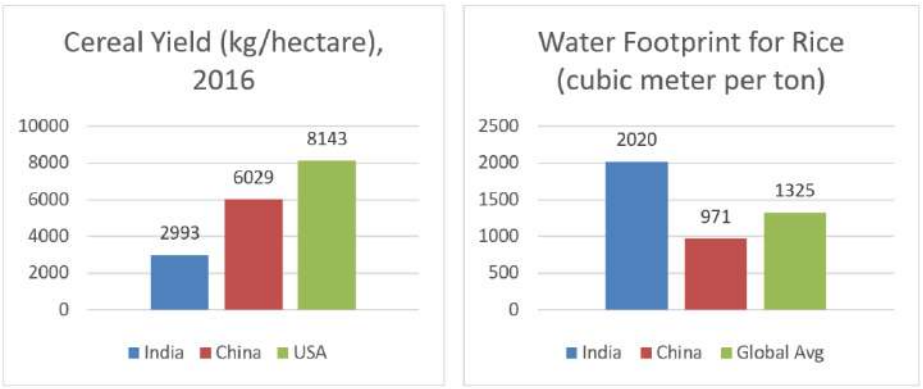
Despite making impressive progress and receiving government attention, the sector continues to be dependent on unpredictable variables, has weak supply chain and low productivity.

India has not been able to completely remove its exploitative dependence on resource intensive agricultural practices. Degradation of land, reduction in soil fertility, increased dependence on inorganic fertilizers for higher production, rapidly dropping water tables and emerging pest resistance are some of the several manifestations of India's unsustainable agricultural practices. As global climate becomes more vulnerable and unpredictable, dependence on unsustainable and resource intensive agriculture will only heighten the risks of food scarcity and agricultural distress.

The sector suffers from poor resource utilisation, with the production quantum and productivity still being quite low. For example, yield of cereals, comprising a major share of food grain production, in terms of magnitude is significantly lower than that of China and the USA. Technology adoption and efficient resource usage in these two countries are far higher, thus resulting in higher yields.

Similarly, use of water in agriculture continues to be high and sub-optimal. The practice of growing water intensive crops, and inefficient water management, makes India a net exporter of water and puts India's long run agronomic sustainability in question. Despite having just one-third of the gross cropped area under irrigation, agriculture consumes 89% of our extracted groundwater.

Figure 11: *Comparison of yield and water footprint*



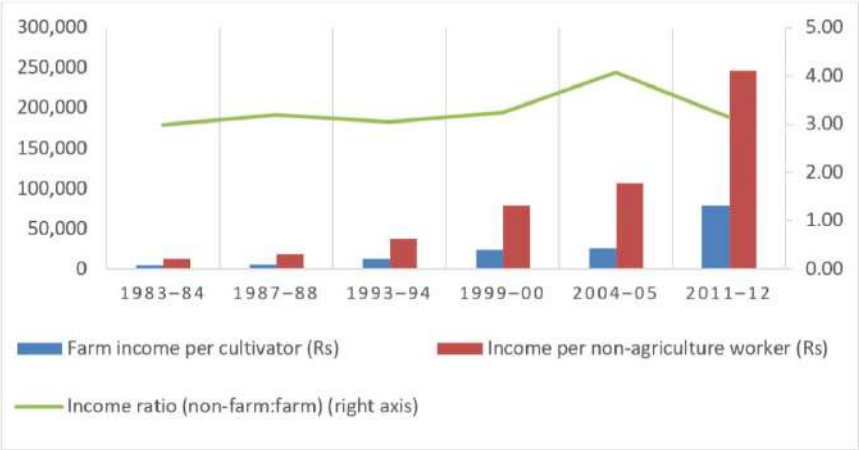
Source: World Bank and UNSECO-IHE Report on Water Footprint 2010

Agrarian distress in India has increased over time due to a multiplicity of factors. Fluctuating agricultural growth rate, globalised value chains leading to variability in commodity prices, unpredictable changes in

[12] *Economic Survey 2018*

monsoon rainfall over years and structurally inefficient domestic agricultural markets are just some of the reasons for income variability of farmers. Various National Sample Survey rounds have shown the reduction in proportion of value share of crops to overall agricultural value from 78% to 69% since the Green Revolution[13]. Thus income disparity between a farmer and non-agricultural worker has increased over the years.[14]

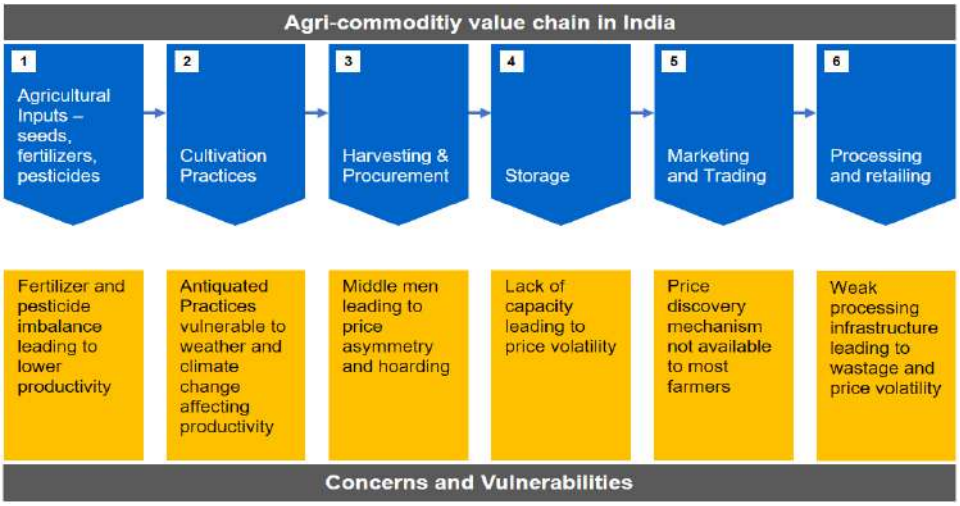
Figure 12: *Income disparity*



Source: Doubling Farmers Income Committee Estimates

On the market side, non-existent functional end-to-end agriculture value chains have caused the price realisation for farmers to remain low. Access to, and timely availability of services, across agricultural value chain at the farmers' end thus becomes a challenge. At present, there is no functional mechanism to track the capacity of storage facilities available to the farmer. Value chain is not integrated through its entire length – procurement to market, including ICT, and banking services. The following figure effectively captures the present scenario.

Figure 13: *Agri-Commodity value chain in India*



[13] *Doubling Farmers Income Committee Estimates*

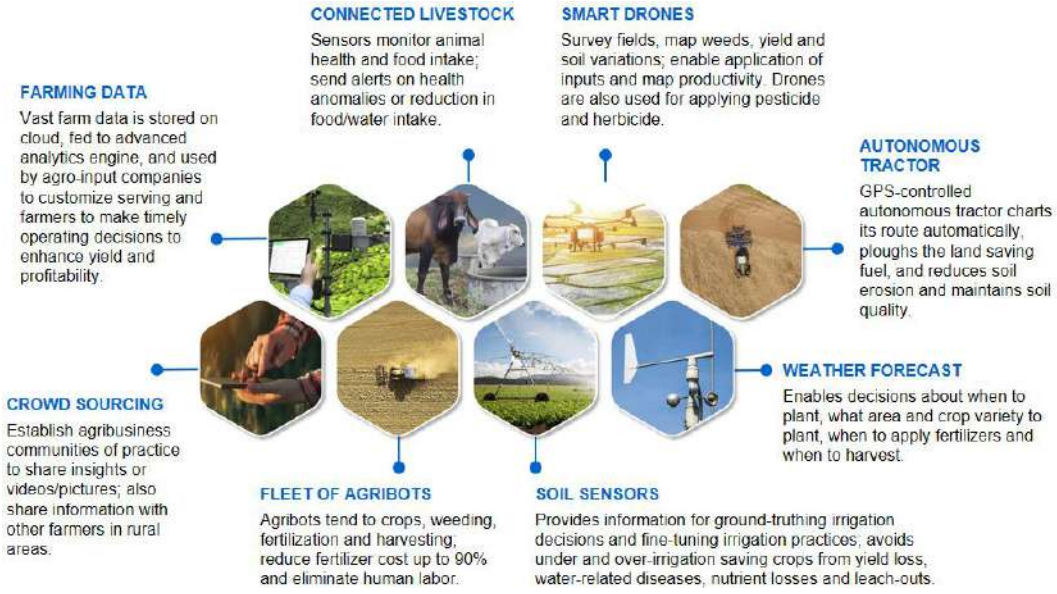
[14] *Chand, R., R. Saxena and S. Rana; Estimates and analysis of farm income in India (2015)*

AI will have significant global impact on agricultural productivity at all levels of the value chain. An estimate by Markets and Markets Research valued AI in agriculture to be USD432 million in 2016 and expects it to grow at the rate of 22.5% CAGR to be valued at USD2.6 billion by 2025[15].

According to CB Insights, agricultural tech startups have raised over USD800million in the last 5 years. Deals for startups using robotics and machine learning to solve problems in agriculture started gaining momentum in 2014, in line with the rising interest in AI across multiple industries like healthcare, finance, and commerce. From analysing millions of satellite images to finding healthy strains of plant microbiome, these startups have raised over USD500 million to bring AI and robotics to agriculture.

Globally, digital and AI technologies are helping solve pressing issues across the agriculture value chain. The relative role of each technology in creating impact is dependent on the nature of the work, and the issues at hand. India has ~30 million farmers who own smartphones, which is expected to grow 3 times by 2020 and 315 million rural Indians will be using internet by 2020[16]. An Accenture study says – *digital farming and connected farm services can impact 70 million Indian farmers in 2020, adding USD9 billion to farmer incomes*. These are not futuristic scenarios, they are in play today, enabled by a vast digital ecosystem which includes traditional Original Equipment Manufacturers (OEM), software and services companies, cloud providers, open source platforms, startups, R&D institutions and others. Future growth is interdependent on the close partnership among these players.

Figure 14: *Ecosystem crucial for benefits of Precision Agriculture*



Source: Accenture Research

In 2016, approximately 50 Indian agricultural, technology based startups ('AgTech') raised USD313 million[17]. For the first time, this sector is seeing widespread participation by startups. Intello Labs, for example, uses image-recognition software to monitor crops and predict farm yields. Aibono uses agri-data science and AI to provide solutions to stabilise crop yields. Trithi Robotics uses drone technology to allow farmers to monitor crops in real time and provide precise analysis of their soil. SatSure, a startup

[15] *MarketsAndMarkets*

[16] *Forbes.com "For India's Farmers It's Agtech Startups, Not Government, That Is Key"*

[17] *Agfunder.com*

with roots in India, uses ML techniques to assess images of farms and predict economic value of their future yield.

Use of AI and related technologies have the potential to impact productivity and efficiency at all of the above stages of the agricultural value chain.

- *Soil health monitoring and restoration:* Image recognition and deep learning models have enabled distributed soil health monitoring without the need of laboratory testing infrastructure. AI solutions integrated with data signals from remote satellites, as well as local image capture in the farm, have made it possible for farmers to take immediate actions to restore soil health.

Box 5: Application for soil care

Berlin-based agricultural tech startup PEAT has developed a deep learning application called Plantix that reportedly identifies potential defects and nutrient deficiencies in the soil. The analysis is conducted by software algorithms which correlate particular foliage patterns with certain soil defects, plant pests and diseases. The image recognition app identifies possible defects through images captured by the user's smartphone camera. Users are then provided with soil restoration techniques, tips and other possible solutions.

- *Crop health monitoring and providing real time action advisories to farmers:* The Indian agriculture sector is vulnerable to climate change due to being rain dependent. Varying weather patterns such as increase in temperature, changes in precipitation levels, and ground water density, can affect farmers especially in the rainfed areas of the country. AI can be used to predict advisories for sowing, pest control, input control can help in ensuring increased income and providing stability for the agricultural community. For example, many agronomic factors (such as vegetation health and soil moisture) can be monitored up to the farm level through remote sensing. Using remote sensed data, high resolution weather data, AI technologies, and AI platform, it is possible to monitor crops holistically and provide additional insights to the extension workers/farmers for their farms as & when required.

- *Increasing efficiency of farm mechanisation:* Image classification tools combined with remote and local sensed data can bring a revolutionary change in utilisation and efficiency of farm machinery, in areas of weed removal, early disease identification, produce harvesting and grading. Horticultural practices require a lot of monitoring at all levels of plant growth and AI tools provide round the clock monitoring of these high value products.

Box 6: AI sowing app

Microsoft in collaboration with ICRISAT, developed an AI Sowing App powered by Microsoft Cortana Intelligence Suite including Machine Learning and Power BI. The app sends sowing advisories to participating farmers on the optimal date to sow. The best part – the farmers don't need to install any sensors in their fields or incur any capital expenditure. All they needed was a feature phone capable of receiving text messages. The advisories contained essential information including the optimal sowing date, soil test based fertilizer application, farm yard manure application, seed treatment, optimum sowing depth, and more. In tandem with the app, a personalised village advisory dashboard provided important

insights into soil health, recommended fertilizer, and seven-day weather forecasts. In 2017, the program was expanded to touch more than 3,000 farmers across the states of Andhra Pradesh and Karnataka during the Kharif crop cycle (rainy season) for a host of crops including groundnut, ragi, maize, rice and cotton, among others. The increase in yield ranged from 10% to 30% across crops.

Box 7: AI for herbicide optimisation

Blue River Technology has designed and integrated computer vision and machine learning technology that enables farmers to reduce the use of herbicides by spraying only where weeds are present, optimising the use of inputs in farming – a key objective of precision agriculture.

- *Increasing the share of price realisation to producers:* Current low levels of price realisation to farmers (as low as 20% in fruits and vegetables[18]) are primarily due to ineffective price discovery and dissemination mechanisms, supply chain intermediary inefficiency and local regulations. Predictive analytics using AI tools can bring more accurate supply and demand information to farmers, thus reducing information asymmetry between farmers and intermediaries. As commodity prices are interlinked globally, big data analysis becomes imperative. Data from e-NAM, Agricultural Census (with data on over 138 million operational holdings), AGMARKET and over 110 million Soil Health Samples provide the volumes required for any predictive modelling.

Box 8: AI for Precision Farming

NITI Aayog and IBM have partnered to develop a crop yield prediction model using AI to provide real time advisory to farmers. IBM's AI model for predictive insights to improve crop productivity, soil yield, control agricultural inputs and early warning on pest/disease outbreak will use data from remote sensing (ISRO), soil health cards, IMD's weather prediction and soil moisture/temperature, crop phenology etc. to give accurate prescriptions to farmers. The project is being implemented in 10 Aspirational Districts across the States of Assam, Bihar, Jharkhand, Madhya Pradesh, Maharashtra, Rajasthan and Uttar Pradesh.

[18] *DFI Committee Report*

Education

An effective education sector has the ability to transform a country through development of human resources and increased productivity. In the context of emerging countries particularly, levels of education and literacy of the population play an important role in development and overall transition to an advanced economy.

In India, the importance of a developed education sector is amplified by a large youth population. Estimates indicate that currently over half the population of the country is below the age of 25. As the adoption of digital means of gathering data increases, it is important that these methods are effectively leveraged to deliver improved education and teaching.

The adoption of technology in education is improving, though not at the pace required. It is estimated that schools globally spent nearly USD160 billion on education technology, or 'EdTech', in 2016, and forecast spending to grow 17% annually through 2020. Private investment in educational technology, broadly defined as the use of computers or other technology to enhance teaching, grew 32% annually from 2011 through 2015, rising to USD4.5 billion globally. Adoption of new technologies is still lacking, however, often attributed to unwillingness of teachers and students to adopt technology.

School education in India has seen substantial progress in recent decades, with efforts at both the Central and State levels, and substantive gains in enrolment have been achieved – Gross Enrolment Ratio (GER) is 97% at elementary level and 80% at secondary level, as per recent figures. However, low retention rates and poor learning outcomes mar the impact of gains in enrolment.

a) *Low retention rates:* Enrolment of children is of little use if children are not retained in the schooling system. Retention rate of 70.7% at elementary level indicates that one-third of enrolled children drop out before completing Class 8. Retention rate at secondary level is also poor at 57.4%. Low quality of education is one of the causes of poor retention.

b) *Poor learning outcomes:* There is increasing concern about the poor learning levels of children in school, and a new National Achievement Survey (NAS) was recently conducted in November 2017. Previous rounds of NAS results provide an insight into longitudinal performance over time – average performance of States / UTs on previous rounds showed that over 60% of Class 5 students scored below 50% across subjects; and for majority of the 31 States / UTs tested, performance significantly deteriorated in NAS Cycle-4 versus Cycle-3. Assessments from the perspective of basic foundational skills also indicate poor learning outcomes – in rural areas, only 47.8% of Class 5 children could read Class 2 level text and only 26% could do Class-5-level arithmetic.

The above scenario is a consequence of a complex interplay of factors that pose challenges to improving the quality of education:

a) *Multi-grade and multi-level classrooms:* For a large proportion of schools, especially in small or remote villages, it is not viable to have separate classrooms and teachers for different grades / classes. Consequently, the teacher is faced with a heterogenous group of children in the same classroom, with wide variations in their classes, ages, abilities and learning levels. This large variation poses a huge challenge to the teacher and is a common cause of poor teaching-learning, thus leading to poor learning outcomes.

c) *Lack of interactive pedagogy and ineffective remedial instruction:* Teaching-learning processes in most classrooms are highly rote-based and non-interactive. Remedial instruction, where conducted, typically lacks customisation to the child's learning level, abilities, and pace of learning.

d) *Inadequate attention / action for likely drop-outs:* Several children may be at risk of dropout due to various factors, such as inadequate school infrastructure, poor teachers, poor school readiness, language barriers, large learning gaps with respect to grade level, family circumstances (e.g. migrant families), poor nutritional or health status, etc.

e) *Large teacher vacancies due to uneven distribution across locations:* Large number of teacher vacancies are mostly not due to an overall shortage of teachers in a State – instead, they are due to uneven distribution across different geographical areas within the State. For instance, recent figures for Uttar Pradesh revealed 1.74 lakh teacher vacancies at elementary school level, but a simultaneous surplus of 0.66 lakh teachers across the state.

f) *Professional development courses / training do not cater to real needs and have poor coverage* Existing teacher training is typically a generic kind of an exercise. It is not linked to the specific weaknesses / requirements of a teacher – for instance, a teacher with poor arithmetic understanding requires corresponding training to clarify arithmetic concepts. Consequently, most teacher training exercises end up as wasted public expenditure, with little or no benefit to the teacher and her / his students. Similar issues exist with respect to training of other staff such as school headmasters/principals. The coverage of existing training programs is also extremely low, typically less than 20% annually.

g) *Low adoption of existing technologies*: A recent survey found that level of adoption of technology in schools is lacking, and can be largely attributed to lack of teacher training, despite provision of the ICT infrastructure. While 83% of the teachers surveyed use computers, the use is limited primarily to audio / visual display, or student practice. A meagre 41% and 27% use technology for tracking student data and participating in forums respectively. This trend is even more pronounced in the low fee school segment surveyed[19]. Another trend observed is that trained teachers are much more likely to use technology in the classroom. 88% of trained teachers reported making use of available computers as compared to only 53% of untrained teachers. Trained teachers were found to be nearly twice as likely to report using technology for communication purposes and for online forum participation[20].

According to EdTechXGlobal, EdTech is becoming a global phenomenon, and as distribution and platforms scale internationally, the market is projected to grow at 17.0% per annum, to USD252 billion by 2020. India's digital learning market was valued at USD2 billion in 2016 and is projected to grow at a CAGR of 30%, reaching USD5.7 billion in 2020 as per estimates from Technopak.

As per Forbes, in 2017, across every market involved in EdTech, international funding reached a new record of USD9.52 billion, and 813 different EdTech companies received fundings last year. These EdTech investments mark a gain of 30% from 2016. VC interest in the education space continues to grow. For example, one of India's leading EdTech startups Byju's raised USD40 million from Tencent in July 2017, just four months after raising USD30 million from Belgium-based Verlinvest. Among Byju's other investors include Sequoia Capital and The Chan Zuckerberg Foundation.

AI has the potential to bring about changes in the sector by supplementing pedagogy and establishing systems to inform and support decision making across stakeholders and administrative levels. However,

[19] *Cross Square Foundation, Ed Tech Adoption Survey*

[20] *Economic Survey, 2016-17*

implementation of AI must be preceded by efforts to digitise records of teacher performance, student performance, and curriculum. Several AI tools are being successfully used in other parts of the world, and they can be adapted to the Indian context to target specific challenges.

a) *Adaptive learning tools for customised learning*: While AI may not completely replace a teacher, it has the potential to greatly assist teachers in efficiently and effectively managing multi-level / multi-grade classrooms, by judging learning levels of individual students, and allowing automated development of customised educational content adapted to each child's class and learning level. Assessing time spent by a student on each part / page of the learning material, for example, would allow real-time feedback on student performance to help the teacher appropriately tailor her guidance to the child. This concept can be extended to automatic grading of tests, as well.

b) *Intelligent and interactive tutoring systems*: Intelligent Tutoring Systems can provide great benefit to students through delivery of learning materials adapted to the child's proficiency level, learning style, and pace of learning. In-built pop-up questions tailored to students, for example, can help increase interactivity, and catch student's attention and interest. It can also help in assessment of student's level of attention or comprehension to appropriately design remedial instruction. GradeGuardian, for example, uses predictive models and visualisations for student performance with an interactive dashboard showing anticipated effect of policy changes. Submission includes 3 components packaged as a single web app – a Chatbot that inputs student information, an Advisor Console that shows students at risk, and a prediction module for policymakers.

Box 9: Creating 'smart content' for improved interactivity

Content Technologies Inc. (CTI), an AI research and development company, develops AI that creates customised educational content. Using deep learning to absorb and analyse existing course materials, textbooks, and course curriculum, the technology creates custom learning materials, including textbooks, chapter summaries, and multiple-choice tests.

A recent hackathon conducted by NITI Aayog also featured 'ReadEx', an android application that does real-time question generation using NLP, content recommendations, and flashcard creation.

c) *Predictive tools to inform pre-emptive action for students predicted to drop out of school:* Analysis of test results and attendance records using AI can be used to predict probable student activities and inform pre-emptive action. For instance, in a recent preliminary experiment conducted in Andhra Pradesh, AI applications processed data on all students based on parameters such as gender, socio-economic factors, academic performance, school infrastructure, teacher skills, etc., with the objective of helping the government identify students likely to drop out. Test results could inform suggestions to enroll students in vocational studies. Additionally, redressal mechanisms could be put in place to identify students whose performance can be improved by focus of existing schemes to their family.

Box 10: Microsoft is helping in predicting drop outs in Andhra Pradesh

The AP government is making concerted efforts to bring down the school dropout rate in the state. It has tied up with Microsoft to address this complex challenge. Based on specific parameters, such as gender, socio-economic demographics, academic performance, school infrastructure and teacher skills, an application powered by Azure Machine Learning processes the data pertaining to all students to find predictive patterns.

With these data insights, the district education officials can intervene and help students who are most likely to drop out. A variety of programs and counselling sessions could be conducted for these students and their parents.

The Andhra Pradesh government, based on machine learning and analytics, has identified about 19,500 probable dropouts from government schools in Visakhapatnam district for the next academic year (2018-19).

d) *Automated rationalisation of teachers:* AI tools can be used to develop automated teacher posting and transfer systems, using analytics based on demand – supply gaps across schools in the State, candidate's prior postings, candidate preferences, etc. This would help in plugging of gaps in teacher distribution more effectively.

e) *Customised professional development courses:* To tackle issues of poorly designed professional development courses with poor coverage, adaptive AI tools can be used to design automated, customised professional development training content for the teacher based on their performance, identification of their knowledge and skill gaps. This could then be continuously adapted as teacher's skills and concepts improve.

Box 11: WriteToLearn by Pearson

Pearson's WriteToLearn software uses natural language processing technology to give students personalised feedback, hints, and tips to improve their writing skills. In describing his experience using WriteToLearn, one 7th-grade English language arts teacher said, "*I feel it's pretty accurate. … Is it perfect? No. But when I reach that 67th essay, I'm not [really] accurate, either. As a team, [WriteToLearn and I] are pretty good*." Essay grading technology cannot substitute for a teacher's ability to provide feedback and coaching on particular words and sentences: the software merely rates students' essays in general areas— such as organisation, idea development, and style—and then provides generic suggestions for improvement in these areas. But when teachers use the software as a first pass at grading and then interject their detailed feedback to address the improvement areas identified by the software, essay grading becomes a much less time-consuming and laborious process. The net result is that teachers can spend less time grading and more time teaching, while also giving students more opportunities to receive customised feedback on their writing.

Smart Cities and Infrastructure

India is currently in the midst of a surge of urbanisation. While the percentage of the population living in urban areas was estimated to be 31% in 2011[21], recent research on satellite data indicates that this figure is close 45% today[22], and predicted to rise to up-to 60 percent by 2050[23]. Though seen as an important aspect of a country's economic growth and a major step in the overall development of the country, *unplanned* urbanisation presents challenges such as congestion, over pollution, high crime rates, poor living standards, and can potentially put a huge burden on the infrastructure and administrative needs of existing Indian cities.

To tackle these challenges, the Government of India has embarked on an ambitious initiative to set up Smart Cities across India, aimed at driving economic growth and improving the quality of life, by harnessing IT solutions. As part of the Smart Cities Mission, 99 cities have been selected with expected investment of INR2.04 lakh crores. The strategic components of these Smart Cities include city improvement (retrofitting), city renewal (redevelopment) and city extension (greenfield development) in addition to a pan-city initiative in which smart solutions are applied covering large parts of the city. The Atal Mission for Rejuvenation and Urban Transformation (AMRUT) is another related initiative which targets improving the infrastructure of existing cities.

Smart cities attempt to address the challenges of urbanisation through development of features based on IT solutions, some of which are listed below.

a) *Poor urban planning*[24]: Smart cities aim to solve challenges of inefficient land use, improper land use categorisation, area based development and lack of open spaces such as parks, playgrounds, and recreational spaces in order to enhance the quality of life of citizens, reduce the urban heat effect, and generally promote improved ecological balance.

b) *Inefficient utility distribution*: Through large scale deployment of smart meters in both electricity and water, smart cities being developed are trying to solve challenges such as low visibility on usage of utilities such as electricity, water, and waste management. This is also targeted to help address issues of leakages in electricity and water distribution, and improper disposal of waste, and have the potential to significantly reduce cost associate with administration and management.

c) *Improved delivery of citizen services:* In the domain of service delivery, smart cities aim to harness data to solve issues in low accountability and transparency. By using digital channels, they can help address challenges in administration of offices, and long wait times. Today, poor standards of grievance redressal form another pressing issue that may be addressed by increasing adoption of technology based solutions.

d) *Improving public safety:* Cities in India today are hotbeds for a range of crimes. Smart cities aim to address the issues of increase in crime and increased risk of urban emergencies through improved city design and surveillance analytics.

Some Smart Cities have already begun implementing these features through specific projects. Pune, for example, has launched The Pune Street Light Project to setup energy efficient street lights that can be remote controlled through a Supervisory Control and Data Acquisition (SCADA) systems. Surat has built a network of more than 600 surveillance cameras which will be expanded to all major locations in the

[21] *Census 2011*

[22] *LiveMint: "How much of India is actually urban?"*

[23] *LiveMint: "60% of India's population to live in cities by 2050: government"*

[24] *Smartcities.gov.in*

city, as well as collaborated with Microsoft to develop solutions for water management and urban planning.

Due to the large amount of data they can create, smart cities are especially amenable to application of AI, which can make sense of the data being generated, and transform it into predictive intelligence – thus transitioning from a smart city to an 'intelligent city'.

However, the wide range of connected devices also gives rise to increased risks in cyber security, with harmful actors such as hackers now capable of affecting city scale infrastructure.

Some use cases of AI that can augment the features of a smart city are listed below.

a) *Smart Parks and public facilities*: Public facilities such as parks and other spaces contribute substantially to a city's liveability. Use of AI to monitor patronage and accordingly control associated systems such as pavement lighting, park maintenance and other operational conditions could lead to cost savings while also improving safety and accessibility.

b) *Smart Homes*: Smart homes concept is creating buzz with AI technologies being developed to optimise human effort in performing daily activities. Extending this concept to other domestic applications such as smart rooftops, water saving applications optimising domestic water utilisation for different human activities etc.

c) *AI driven service delivery*: Implementation of AI to leverage data on service delivery could see application such as predictive service delivery on the basis of citizen data, rationalisation of administrative personnel on the basis of predicted service demand and migration trend analysis, and AI based grievance redressal through chat-bots.

d) *Crowd management:* Use of AI in providing effective solutions in crowd management in recent times have been in vogue and given fruitful results in averting city-scale challenges such as managing mega footfall events, emergency and disasters. Accenture worked with the Singapore Government during their SG50 Celebrations (50th anniversary of Singapore' independence), and developed solution aimed at predicting crowd behavior and potential responses to incidents. The solution resulted in 85% accuracy in high crowd activity, crowd size estimation and object detection. Closer home, the "Kumbh Mela Experiment" is aimed at predicting crowd behavior and possibility of a stampede. Similar Big Data and AI solutions could help with advance prediction and response management.

e) *Intelligent safety systems*: AI technology could provide safety through smart command centres with sophisticated surveillance systems that could keep checks on people's movement, potential crime incidents, and general security of the residents. Social media intelligence platforms can provide aid to public safety by gathering information from social media and predicting potential activities that could disrupt public peace. In the city of Surat, the crime rate has declined by 27% after the implementation of AI powered safety systems.

f) *Cyber*-attacks: Cyber-attacks seem to pose a great threat to our institutions and public systems, today. AI technologies possess the capability to detect vulnerabilities and take remedial measures to minimise exposure of secure online platforms containing highly sensitive data from being targeted by unscrupulous social elements.

Smart Mobility and Transportation

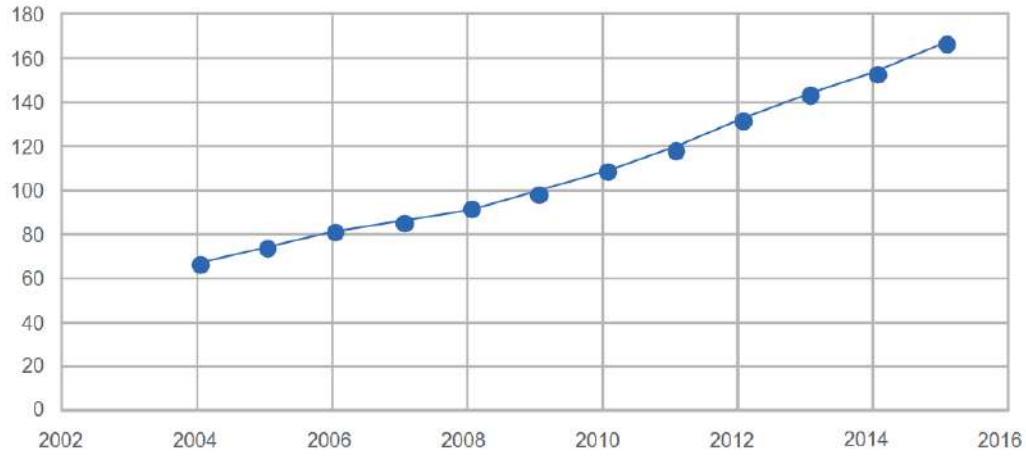
Mobility and transportation form the backbone of the modern economy due to their linkages with other sectors and importance in both domestic and international trades. Today's society demands a high degree of mobility of various kinds, so as to enable efficient and safe transportation of both people and goods. As a major contributor to overall emissions, this sector must also be sensitive to ideas of environmental sustainability.

In India, majority of both passenger and freight traffic is carried through roads and railways. As of 2007 – 08, roads and railways accounted for almost 87% of total freight traffic in the country and almost 90% of total traffic as of 2011-12[25]. As the economy grows, it is expected that this reliance on these modes of transport shall continue unless there are major shifts in the policy initiatives in the area. The fact that these modes of transport are particularly pollution intensive compared to shipping and air transport, further increase the need to implement smart practices in their deployment.

Even apart from issues in poor modal mix, the Indian transportation sector faces a variety of issues.

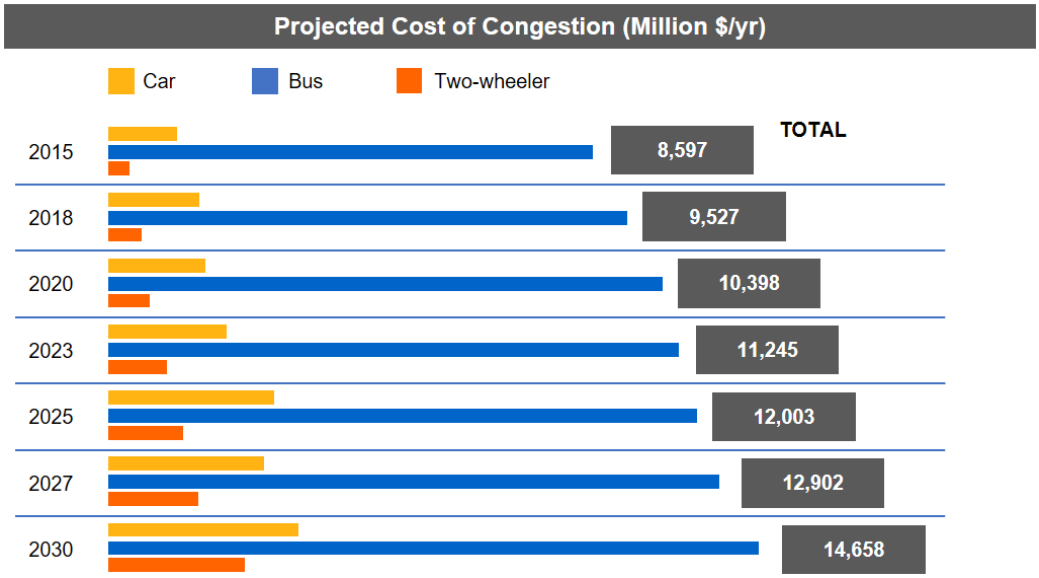
a) *Congestion and road accidents:* Despite having one of the most extensive transportation networks in the world, various sub-sectors sector continues to be underdeveloped leading to economic inefficiency and enormous human cost. Congestion and its associated costs in India are continually on the rise. Statistics from Government of India and a study conducted by IIT Madras suggest the following growth patterns over the years, respectively.

Figure 15: *Total registered motor vehicles per 1,000 population*



[25] *NTDPC, India Transport Report: Moving India to 2032*

Figure 16: Projected cost of congestion (USD million / year)

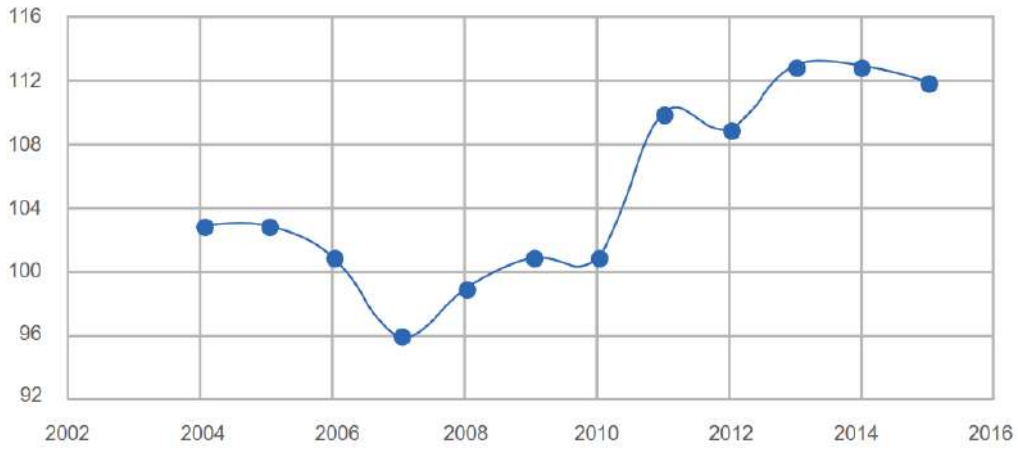


Source: Cost Estimates for Road Congestion in Delhi, IIT (Madras)

b) *High number of traffic deaths:* According to a PIB release by the Ministry of Road Transport and Highways (MORTH) in March 2017, the total number of road accidents in the country during 2015 was 501,423 which resulted to 146,133 fatalities. The National Highways (NHs) accounted for about 29.1% share of total road accidents and 35.0% of total fatalities. Although the existing NHs comprise approximately 1.9 % of total road network, they carry about 40% of total road traffic. According to MORTH statistics, there has been a steady increase in the number of on-road accidents over the years.

c) *Lack of public transportation infrastructure:* Public transport infrastructure development remains laggard in the overall discourse of transport policy design, either at national and regional levels, with focus directed towards promoting and improving private car and associated infrastructure. The following statistics from MORTH indicates the modal share of public transport (buses) over the years which has seen, but minimal increase over the years.

Figure 17: *Total numbers of public buses per 1,000 population*



Source: MORTH

d) *Assisted vehicle technologies*: Autonomy is not economically viable in India currently as driver costs per kilometre is too low. However, investing in the suite of autonomous vehicle technologies and exporting such vehicles represents a significant economic opportunity for India and since the same technologies can play a large role in reducing fatalities and decreasing congestion, it would be wise for Indian manufactures to invest in research and development of the broader suite of technologies that are essential for assistive AI. These technologies can assist the driver by taking driving decisions which the system has a high degree of confidence in and alerting the driver in case it has a low degree of confidence in any decision. A prime example of such productised, assistive AI is the advanced cruise control used in Tesla vehicles today. This can follow highway traffic and the curves in the road as well as start and stop in response to traffic. However, the moment the driver gives any input, the system is taken over by the driver. This hybrid approach is much safer than an unassisted human driving, without the potential drawbacks of having a completely autonomous system. There is another reason why India should not completely ignore assistive vehicle technology research, and it has to do with the development of new public infrastructure. Since we have only recently begun building a large share of the total requirement of greenfield infrastructure, we have the benefit of hindsight.

e) *Need for sustainable transportation*: The recent initiative of the Government of India for announcing development of 100 Smart Cities is aimed at addressing this anomaly and catalyse smart strategies for urban planning which promote sustainable land use design and multimodal integration. While new initiatives could take time to show realisable impact, the existing issues in urban mobility related to congestion, efficient traffic flow, movement of goods etc. can indeed be solved using AI technology. AI can power multimodal integration by assisting with scheduling public transportation systems, improved accessibility to public transportation infrastructure based on users' choice behaviours, while also suggesting real-time travel mode advisory based on predicted traffic situation. AI enabled mobility solutions can ameliorate several of the challenges being faced by the Indian automotive and transportation sector.

f) *Efficiencies in design of greenfield infrastructure* – Autonomous-ready traffic will have significant impacts on greenfield road infrastructure design and consequently on greenfield city design. Lane size, lesser traffic congestion and reduced costs in upgradation of highway infrastructure are some of the externalities which will benefit the sector of assisted vehicle adoption.

Listed below are some of the major applications of AI on the mobility front beyond autonomous cars:

a) *Autonomous trucking:* Autonomous technology in trucking has the potential to transform the way we move goods today. AI can help increase safety and hauling efficiency through intelligent platooning, wherein trucks form platoons giving drivers the liberty to rest while the platoon keeps moving. Such a method also ensures optimal road-space utilisation, helping improve road infrastructure capacity.

b) *Intelligent Transportation Systems:* Through the use of an intelligent traffic management system including sensors, CCTV cameras, automatic number plate recognition cameras, speed detection cameras, signalised pedestrian crossings and stop line violation detection systems and the use of AI, real time dynamic decisions on traffic flows such as lane monitoring, access to exits, toll pricing, allocating right of way to public transport vehicles, enforcing traffic regulations through smart ticketing etc. can be made. Accident heat maps could be generated using accident data and driver behaviour at specific locations on the road network related to topology, road geometric design, speed limit etc. and suitable measures could be pre-emptively taken to prevent possible accidents. Also, AI could help to design sophisticated urban traffic control systems that can optimise signal timings at the intersection, zonal and network level, while also facilitating services such as automatic vehicle detection for extension of red/green phase or providing intermittent priority.

c) *Travel route/flow optimisation:* With access to traffic data at the network level, AI can help make smart predictions for public transport journeys by optimising total journey time including access time, waiting time and travel time. Considering factors such as accessibility to nearest mode of travel, most convenient access path based on local conditions and one's preferences, AI can revolutionise first-last mile travel which could change the way we perceive public transport journeys, today. About private car usage, AI could utilise a range of traffic data sets and one's own preferences to make human-like decisions on route selection. With information on dynamic tolls and traffic flows on links, the dependency on overhead Variable Messaging Systems (VMS) could be minimised, reducing substantial infrastructure costs. On the systemic level, AI can help predict flow of traffic at the network level and suggest alternative flow strategies in order to contain congestion, alleviating cities of this major issue.

d) *AI for Railways:* According to official figures, more than 500 train accidents occurred between 2012-2017, 53% of them due to derailment. Train operators can obtain situational intelligence through real-time operational data and analyse them in three different dimensions: spatial, temporal and nodal. Fleet management and asset maintenance including that of rolling stock are pertinent AI use cases. Recently, the Ministry of Railways, Govt. of India has decided to use AI to undertake remote condition monitoring using non-intrusive sensors for monitoring signals, track circuits, axle counters and their sub-systems of interlocking, power supply systems including the voltage and current levels, relays, timers.

e) *Community Based Parking:* The availability of parking is a major issue for Indian cities. AI can help optimise parking, likely by minimising vehicle downtime and maximising driving time. With the advent of electric vehicles, AI will be needed to mediate the complex vehicle grid interactions(VGI) as well as for charging optimisation. Parking guidance systems help drivers to find vacant parking spaces while they are using the road network and have approached close to their destination. Community

based parking using AI helps cars in traffic to collect data on vacant parking spaces, and allocates cars to spaces such that the demand is always met.

Box 12: Effect on R&D

Globally, research on autonomous vehicle has spurred advances, especially in AI fields of computer vision and robotics. Due to the extremely high market potential, over the past two years, most of the large investments in AI have been made in the field of autonomous vehicle as it is widely tipped to be the first large scale commercial application of AI to be adopted.

Moreover, due to the congestion and chaotic conditions of Indian traffic, AI algorithms trained on Indian driving data have the potential to be very robust.

Error rates of object classification have fallen from 28.5% to 2.5% since 2010 according to the Stanford AI index. Therefore, current techniques are mature enough to be used in Indian conditions. Also, within AI, the core technologies used have high transference potential. The same template used to identify objects on a road can be used to identify cancerous cells in a pathological image.

Key challenges to adoption of AI in India

Commonality of problems mandate an integrated approach

The preceding analysis of focus sectors – Healthcare, Agriculture, Education, Smart Cities and Infrastructure, and Smart Mobility and Transport, highlight the potential of AI tools and technologies in transforming the sectors and state of Indian economy as a whole. The analysis, however, also detail a multitude of challenges that India needs to overcome to realise the full potential of a disruptive technology like AI.

Adopting a narrow view and focusing on the challenges for a specific sector, the barriers to developing a robust set of AI applications may seem contextual and limited to that sector. Taking Healthcare sector as an example, enabling large scale adoption would require at least the following factors to be addressed:

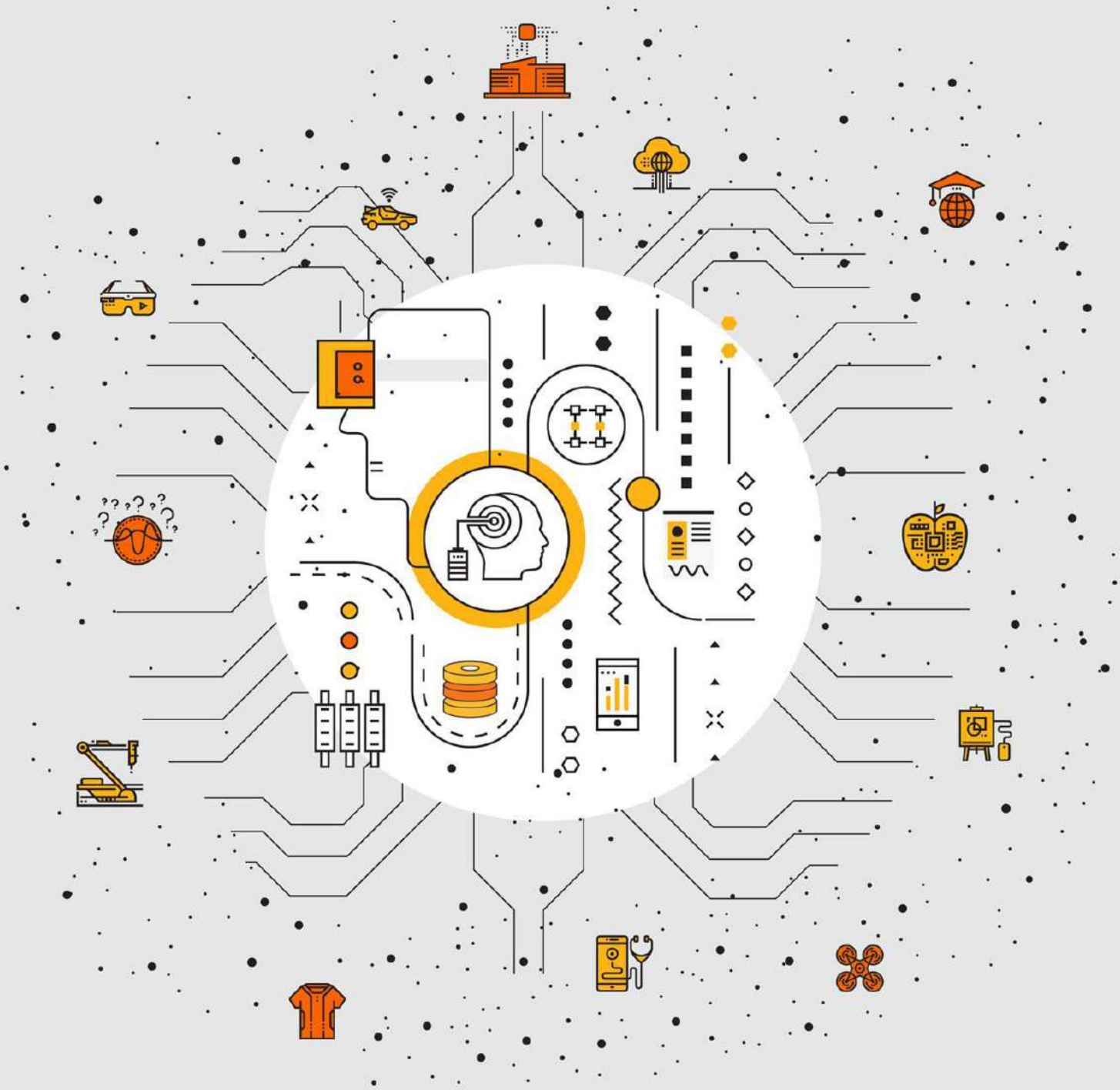
a) absence of collaborative effort between various stakeholders: while India has adopted electronic health record (EHR) policy, sharing of data between various hospital chains still remains a work in progress, since different hospital chains have adopted different interpretations of 'digitising records';
b) relevant data is unavailable and there is absence of robust open clinical data sets; and
c) concerns on privacy and security of data, including lack of formal regulation around anonymisation of data.

However, analysing across the focus sectors, the challenges are concentrated across common themes of:

a) Lack of enabling data ecosystems
b) Low intensity of AI research
    i. Core research in fundamental technologies
    ii. Transforming core research into market applications
c) Inadequate availability of AI expertise, manpower and skilling opportunities
d) High resource cost and low awareness for adopting AI in business processes
e) Unclear privacy, security and ethical regulations
f) Unattractive Intellectual Property regime to incentivise research and adoption of AI

These challenges, while by no means exhaustive, if addressed in an expeditious manner through concerted collaborative efforts by relevant stakeholders, with government playing a leading role, could lead to fundamental building blocks that form the core to India's march towards leadership in AI. The next section of the paper attempts to solve some of these challenge through specific interventions and recommendations. These recommendations have been formulated as fundamentally infrastructural in nature, and hence span across sectoral use cases.

PART III:
RECOMMENDATIONS



Way Forward to Harness the Power of AI

Recommendations

India's unique challenges and aspirations, combined with the advancement in AI, and a desire to assume leadership in this nascent technology means India's approach towards AI strategy has to be balanced for both local needs and greater good. The way forward for India in AI has to factor in our current strengths in AI, or a lack thereof, and thus requires large scale transformational interventions, primarily led by the government, with private sector providing able support.

This section lays down a set of recommendation to address the biggest challenges and opportunities for India in the field of AI. The preceding analysis of focus sectors lead us to the assertion that the efforts need to be concentrated across major themes of research, data democratisation, accelerating adoption and reskilling – with privacy, security, ethics and intellectual property rights permeating as common denominators for all our recommended initiatives. These challenges, while by no means exhaustive, if addressed in an expeditious manner through concerted collaborative efforts by relevant stakeholders, with government playing a catalytic role, could lead to fundamental building blocks that can form the core to India's march towards achieving its goal of #AIforAll.

India's capabilities in AI research are rather limited, both in quantity (distant 5th globally) and especially in quality (disappointing impact of research produced). The research community is rather confined to a handful of academic institutes, and relies on individual brilliance rather than institutional competence. Acerbating the problem is the fact that private sector's contribution to AI research has remained meagre. Despite some encouraging recent developments, viz. Government of Karnataka's intention to set up a Centre of Excellence in AI in partnership with NASSCOM, a lot of ground needs to be covered. The first set of recommendations focus on turbocharging both core and applied research. In addition, two frameworks for solving some of AI's biggest research challenges through collaborative, market oriented approach have been proposed.

The new age of AI and related frontier technologies would disrupt the nature of jobs of tomorrow and the skills required to realise the true potential of these transformative technologies. The changes and challenges anticipated for the workforce will come from both the demand and supply side: demand for capabilities for jobs that don't even exist today and diminished demand for some of the jobs that could be automated, supply of newly minted STEM graduates, a large portion of whom may struggle to be gainfully employed. Given our strength in advanced IT sector and the strength of favorable demographics, India may seem more equipped for workforce disruption that AI will bring, however our large numbers may soon turn from potential assets to liabilities if right structures are not put in place. Our next set of recommendations focus on reskilling of existing workforce and preparing students for developing applied set of skills for the changing world of technology.

Early adoption of AI – be it the research community building technology infrastructure, the startup community developing applications and corporations deploying solutions for their business needs, would be one of the key determinants in ensuring leadership in AI. Adoption of AI in India has remained rather limited, less than a quarter of firms in India are using AI in any form for their business processes and startup ecosystem in AI is virtually non-existent. Among the several impediments towards large scale

adoption of AI in India, the primary ones include difficulty in access to data (more specifically, structured and intelligent data), high cost and low availability of computing infrastructure, lack of collaborative approach to solving for AI combined with low awareness. Our recommendations to address these challenges include developing large foundational annotated data sets to democratise data and multi-stakeholder marketplaces across the AI value chain (data, annotated data and AI models).

One of the key aspects of our ambition of #AIforAll includes responsible AI: ensuring adequate privacy, security and IP related concerns and balancing ethical considerations with need for innovation. Our final set of recommendations lay down the challenges and suggestion for addressing some of these not so straightforward implementational challenges of AI.

The recommendations in the following chapters are aimed at initiating an informed conversation on India's future roadmap for AI, and are descriptive rather than prescriptive by design. The paper should be seen as providing framework for developing National Strategy for Artificial Intelligence, and as such, we have consciously avoided providing specific funding targets and funding mechanisms, as these require broad based stakeholder consultations.

Research

Incentivising Core and Applied research in AI

Advanced research, both core and applied, provides the basis for commercialisation and utilisation of any emerging technology, more so for technologies like AI.

Where does India stand in Artificial Intelligence research?

India has the necessary building blocks to develop a thriving AI research and development ecosystem, viz. availability of highly educated talent pool, world class educational institutes and an illustrious list of top notch IT companies dominating the global IT landscape. Despite these advantages India sees itself lagging considerably in producing world-class research and innovation in most technology fields, more so in AI.

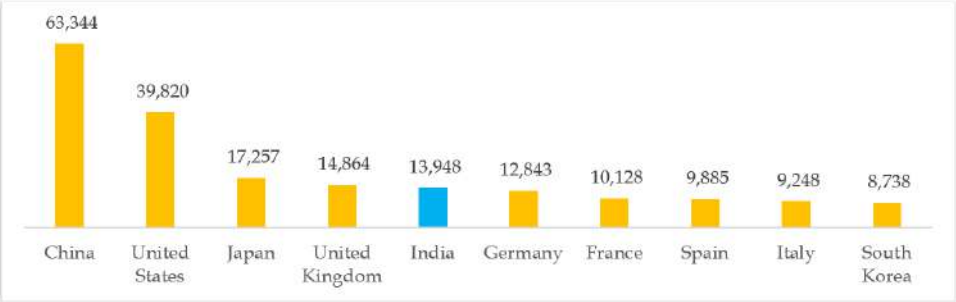
India produced a whopping 2.6 million[26] STEM graduates in 2016, second only to China and more than 4 times the graduates produced by USA, thus producing the requisite talent pool to drive innovation in emerging technologies. Disappointingly though, an overwhelming majority of this talent pool is focused on routine IT development and not so much on research and innovation. Exacerbating the problem further, a majority of the small population focused on research almost always prefers to pursue advance degrees (Masters or PhD degrees) to subsequently apply their expertise abroad.

An analysis of India's competence in core research in AI paints a somber picture. As per the Global AI Talent Report 2018, which crawled LinkedIn for its analysis, India only has 386 of a total of 22,000 PhD-educated researchers worldwide, and is ranked 10th globally. The report also looks at leading AI conferences globally for presenters who could be considered influential experts in their respective field of AI. On this metric, India was ranked 13th globally, with just 44 top-notch presenters. While these two approaches have their limitations and inherent biases, anecdotal evidence based on discussions with top researchers reveals that serious research work in India is limited to less than 50 researchers, concentrated mostly at institutes like IITs, IIITs and IISc.

In terms of the citable documents published in the field of AI from 2010 - 2016, India ranks a distant 5th, far behind the likes of China and USA and just about edging ahead of Germany and France who have considerably smaller STEM population.

[26] *World Economic Forum*

Figure 18: *Citable documents in Artificial Intelligence (2010 - 2016)*



Source: Scimago Journal and Country Rank (SJR)

Diving deeper into these numbers, if we look at the country wise H-Index (a metric that quantifies a country's scientific productivity and scientific impact), India ranks a dismal 19th globally. In other words, while India may be producing research pieces in numbers, their utility has been rather limited.

Figure 19: *H Index for Artificial Intelligence (1996 - 2016)*

Source: Scimago Journal and Country Rank (SJR)

Looking at the research coming out of academic institutes, the numbers are heavily skewed in favour of top-15 institutes, that in total have contributed more than 42% of all research publications from 2001 – 2016[27]. IISc dominates the research publications, with 7.5% of all publications coming from this institute. For a country that has more than 750 universities and close to 40,000 colleges, this concentration of publication is a worrying sign.

The Indian IT services companies, the likes of TCS, Wipro and Infosys, have been the flag bearers of India's competence in implementation of cutting edge technology solutions, yet their contribution to research has been limited. Given that these IT giants have been working closely with businesses globally and anticipating the trends in emerging technologies, it wouldn't be unreasonable to expect a sizeable volume of research work coming out of these companies. Yet, looking at all the research publications from 2001 – 2016, only 14% of all publications have come from industry, with universities contributing 86% of all publications. Even this limited research publication universe by industry is dominated by Indian subsidiaries of international companies (~70%), with only one Indian company in top-10 (TCS)[28].

[27] *Neel Shah: "Research trends of AI in India"*

[28] *Neel Shah: "Research trends of AI in India"*

Box 13: What India can learn from other economies in terms of AI Research

The US Government is estimated to have spent USD1.2 billion in non-classified research in 2016-18 and the Defence Advanced Research Projects Agency (DARPA) is seeking a budget of USD3.44 billion in fiscal year 2019-20, an increase of 8.5% compared with its request for fiscal 2018-19. However, US leadership in AI investment has largely been driven by the private sector. The world's leading companies in AI research in 2016 were Microsoft, Google and IBM, all US companies. According to CB Insights, based on 2017 figures, Amazon, Google and Microsoft dominate enterprise AI – again all US companies. It is estimated that more than half the world's unicorns are from the US. The digital ecosystems around the hubs of Silicon Valley, Seattle, Boston and New York, which bring together talent and research capabilities from leading universities, private investment and cross-science / industry collaboration, can be considered to have played an important role in developing the US's AI capabilities.

While still behind the US in terms of overall investment, China has clear ambitions to be at the same level as the US by 2020 and the world leader in AI by 2030, supported by a new development plan to create a USD150 billion domestic AI industry. Its plans to build a new AI industry include a national fund that supports research, from the most basic research to critical AI projects. The top 9 universities have received government funding to each establish an AI school and the remaining 32 to include an AI programme as part of their curriculum. The Ministry of Industry and Information Technology is planning to put nearly USD950 million dollars per year into strategic AI projects for State Owned Enterprises and the public sector. In addition to state investment, the government is expected (at the time of writing) to publish Next Generation AI Development Guidelines immanently. The guidelines are expected to include a clear governance structure, with allocation of responsibility and plans for research, industry and legislative action or each of 2020, 2025 and 2030.

While China's approach is not necessarily replicable in other parts of the world, there are two key learnings from its programme:

- Public sector investment, particularly in R&D, helps drive private investment.
- It has a plan with a governance structure and clear milestones. Having a plan instils confidence in inward investors. Based on interviews Accenture carried out with inward investors in the UK, there was consensus that the governments' public messaging had a significant impact on companies' confidence and therefore willingness to invest in a country.

In the UK, the universities of Cambridge and Oxford are considered centres of AI innovation; having already stimulated three startups that made major AI breakthroughs and later became prime acquisition targets. Google in 2014 bought DeepMind, Apple in 2015 purchased VocalIQ, and Microsoft bought SwiftKey in 2016. This success is supported by funding from organisations like the Leverhulme Trust, which provides annual funding of GBP80 million for research. Other capabilities include The Alan Turing Institute: the national institute for data science. The Institute was established in 2015 by five founding universities (Cambridge, Edinburgh, Oxford, UCL and Warwick) and the UK Engineering and Physical Sciences Research Council. The Institute's researchers work across disciplines and look at theoretical development and application to real world problems. It was announced as the national centre for AI in November 2017 and six new universities will join the institute 2018.

The German Research Centre for Artificial Intelligence (DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz) is one of the world's largest AI research institutes. It has facilities in the German cities of Kaiserslautern, Saarbrücken, Bremen and Berlin and is partnering with companies on application oriented basic research to develop product functions, prototypes and patentable solutions.

The EU's Robotics Public Private Partnership, launched in 2013, has seen the allocation of EUR700 million for research to 2020. This is coupled with private investment for an overall backing of EUR2.8 billion.
It is believed to be the biggest civilian research programme in this area in the world and could be considered instrumental in the strong presence of Europe among service robot manufacturers. Clearly, public-sector investment has paid-off.

***Source**: Realising the economic and societal potential of responsible AI, Accenture*

The research ecosystem in India has seen some green shoots in recent years. Encouraging is the fact that the number of papers published has jumped 10 fold in last 10 years, from 331 papers in 2006 to 3,301 papers in 2016[29]. IISc, almost all the IITs, some of the IIITs and central / state universities have increased their research efforts in various foundational and applied fields of AI. IIT Bombay and IIT Patna have entered into a joint research collaboration with industry to focus on the applied aspects of AI. The research, focused on IT services and social good, will aim to provide powerful AI insights and recommendations for improved productivity. It also includes software analytics – building, testing, managing and modernisation of applications, solving real-life social issues such as malnutrition, human trafficking and climate change through prediction and recommendation models using AI.

Another research group at IISc is working on the theory and application of reinforcement learning (RL), an aspect of machine learning used in optimisation problems. They are particularly interested in traffic handling—both the vehicular kind on our roads, as well as the digital kind in our wireless networks.

Furthermore, there have been some encouraging efforts from both government and private sector in facilitating top quality research in recent times. Government of Karnataka is setting up a Centre of Excellence for Data Science and Artificial Intelligence in partnership with NASSCOM. Wadhwani Foundation has set up India's first research institute dedicated to developing AI solutions for social good in Mumbai in Feb 2018.

However, the research ecosystem still has several obvious gaps. The Detailed Project Report of Inter-Ministerial National Mission on Interdisciplinary Cyber Physical Systems has highlighted some of these as:

a) *Lack of collaborative / interdisciplinary approach:* research is mostly focused in silos in academic institutions

b) *Lack of scale for experimental validation:* due to various practical and financial reasons, university research is largely restricted to theoretical or laboratory scale. This needs to be augmented with pilot projects / large scale test beds / laboratories

c) *Lack of facilities to support large scale experimental test beds:* Large scale experimental test-beds are difficult to construct, maintain and operate, solely by academic institutions

d) *Lack of connect with stakeholders and practitioners to convert outputs to outcomes:* The views of the stakeholders in terms of what application problems to focus on will be of great importance to ensure practical applicability of the research. At the same time, this should be facilitated in a way which does not constraint/suffocate the academic researchers in order for them to make foundational advances. Involving "technology translators" at an early stage, i.e. entrepreneurs/agencies/companies which can convert the research technologies to commercial products is needed.

[29] *Scimago Journal and Country Rank (SJR)*

e) *Lack of large scale mission mode project management capabilities:* Academic researchers usually work best individually (with a small team of students and research project staff). Current approaches to research and related facilities may not be suited for large scale experimental projects.

While the numbers point to a small yet encouraging base, a concerted effort is needed to build a comprehensive research focused AI strategy for India, one that will position India towards global leadership in this emerging area of technology. What is evident though is that incremental changes would not suffice and there is need for transformational changes to boost research major push coming from the government.

Framework for promoting Artificial Intelligence Research in India

The Detailed Project Report of Inter-Ministerial National Mission on Interdisciplinary Cyber Physical Systems (IM-ICPS) has suggested the following four-tier framework for promoting research focused on all aspects of technology life-cycle: research, technology deployment, translation and management:

a) *ICON (International Centres of New Knowledge):* focusing on creation of new knowledge through basic research,

b) *CROSS (Centre for Research On Sub-Systems):* focusing on developing and integrating core technologies developed at ICON layer and any other sources

c) *CASTLE (Center for Advanced Studies, Translational research and Leadership):* focusing on development and deployment of application based research and

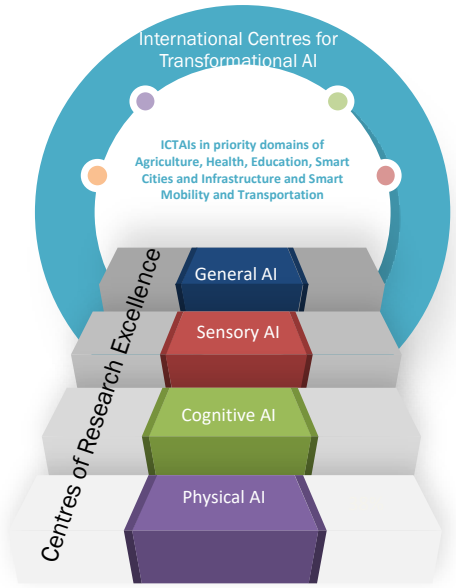
d) *CETIT (Centre of Excellence in Technology Innovation and Transfer):* focusing on commercialisation of technologies developed

While the above structure has significant advantages, a far more simplified and agile approach is required to ensure seamless, targeted and accountable framework for promoting research. Hence the following two-tier integrated approach to boost both core and applied research in AI is proposed:

a) *COREs (Centres of Research Excellence in Artificial Intelligence):* COREs will focus on core research of AI, and will take on the mantle of executing the responsibilities of both ICON and CROSS as per the IM-ICPS framework. Thus, COREs will specialise in creating new knowledge through basic research and will source for fundamental knowledge / technologies that will be needed to keep India prepared for the next generation of technologies. Furthermore, COREs will also emphasise on development infrastructure tools for direct application of basic research, including development of new areas of AI architecture / platforms.

b) *ICTAI (International Centre for Transformational Artificial Intelligence):* ICTAIs will provide the eco-system for application based technology development and deployment, and will take on the mantle of executing the responsibilities of both CASTLE and CETIT as per the IM-ICPS framework. This will be an industry-led initiative and expected to take on the top-level challenges identified or inter-ministerial projects calling for AI based solutions. Furthermore, ICTAIs will also be responsible for delivering commercial technology, and taking ideas / concepts or prototypes and turning them into marketable products by way of proactive coordination, communication and interfacing for technology transfer to the industry.

Figure 20: Proposed integration of COREs and ICTAIs



In summary, COREs will be focused on core research, in evolving and new areas of AI, and will act as the technology feeders for ICTAIs which will be focused on creating AI based applications for, and accelerating early adoption in, domains of societal importance.

In addition, an umbrella organisation should be established to address issues relating to access to finance, social sustainability and the global competitiveness of the technologies developed. This body, which shall be recognised as the Centre for Studies on Technological Sustainability (CSTS), could be established on the lines of the Campus for Research Excellence and Technological Enterprise (CREATE), Singapore program or Innovate UK. The major responsibilities of CSTS could be on the following lines:

a) to monitor the impact of the AI technologies developed at the consumer level through social indices and recommend necessary modifications for a better market penetration

b) to study the financial viability of the AI technologies developed such that it caters to the target consumer base, while proposing improved pricing models for a pan India reach

c) to study best practices on pricing models and social penetration of AI technologies across the world and recommend strategies to foster globally competitive technological development

d) to catalyse international collaborations for COREs and ICTAIs

e) to study the AI landscape in other nations and design strategies for customisation and deployment of developed AI technologies as per their specifications for a global impact

f) to provide a knowledge management platform for AI technologies by organising international workshops and conferences, promoting the confluence of thought leaders, practitioners and authorities

Centres of Excellence for Artificial Intelligence

COREs for AI will focus on core research, in evolving and new areas of AI.

To start with, COREs could be established at IISc, ISI and top IITs and IIITs. Given that the research in AI needs to be multi-disciplinary, linkages need to be established with premier institutions in other disciplines viz. AIIMS for healthcare, TISS for arts and social science etc. Furthermore, these COREs should also act as a guide and mentor for other institutes researching in AI, in a hub-and-spoke model, to enable broad based development of AI research capabilities across India.

The COREs would have to build on both the short-term and the long-term capabilities of these research centers. In the short-term, given the paucity of quality faculty in India in AI, appropriate incentivisation mechanism (which could be a combination of promise of topnotch infrastructural facilities and remunerations matching international standards) to bring top-tier international faculty, especially the Indian diaspora, needs to be developed. Furthermore, the top Indian PhD aspirants, who would otherwise pursue their studies from top universities abroad, will need to be retained in India, again through appropriate incentivisation mechanism. One possible way could be to institute National AI Fellowships.

In the long-term, successive PhD classes of these COREs can increase the faculty pool and work towards a sustainable operational model for COREs.

Possible focus area for the COREs for AI could be:

a) Sensory AI (Computer Vision, IoT etc.),
b) Physical AI (Robotics, Industrial Automation etc.),
c) Cognitive AI (NLP, worker training etc.),
d) General AI,
e) High precision learning from small data sets,
f) Research on new algorithms (e.g. advance cryptography, security), data sets etc., and
g) Explainable AI

An application based model may be followed to selected COREs wherein an applicant institute would have to demonstrate a viability plan, in terms of faculty and other capabilities, before being anointed an CORE. A CORE can choose to work on one or multiple focus areas. The COREs will be encouraged to pursue projects across other COREs, to promote linkages and cross-functional technologies. The financial component of COREs for AI, which could be to the tune of INR 50 crore – INR 100 crore per CORE, should also include large scale funding for specific projects.

International Centers for Transformational Artificial Intelligence

International Centers for Transformational Artificial Intelligence (ICTAIs) are envisioned as institutions focused on creating AI based applications for, and accelerating early adoptions in, domains for societal importance. These applied research and development institutions should be set up with the elements of domain or industry "pull" baked into their structure and operation. While core (or fundamental AI) research and teaching may not be seen as their priority areas, some flexibility might be built in the structure.

From both funding and operational perspective, the ICTAIs are envisioned to be a truly public private partnership. A professionally managed society / section 8 company ("ICTAI Inc.") should be set up and trusted with the initial contribution from the government. The management team for ICTAI Inc. should comprise of suitable representation from the government, but should have a majority of independent private sector representation. The mandate for ICTAI Inc. would be to select the ICTAIs to fund and oversee their progress. In addition, ICTAIs may commission "*moonshot projects*" that may span across multiple ICTAIs or may need limited term independent project teams to be set up.

ICTAIs are expected to leverage the strength of AI technologies to solve for application in a specific focus sector areas. Suggested focus sector areas to begin with, as identified in this report, are Healthcare,

Agriculture, Education, Smart Mobility and Transportation, Smart Cities and Infrastructure. ICTAI Inc. however should have the autonomy to decide on sectors to focus on and sequencing of sector selection. ICTAI Inc. should work proactively with private sector institutions to seek partnerships to set up the ICTAIs in priority domains.

Furthermore, ICTAIs should have well-defined linkages with the COREs. As a technology feeder for ICTAIs, COREs should be incentivised to commercialise their research at ICTAIs.

The key to success for the ICTAIs would be their leadership teams. The decision on which ICTAIs to invest in could involve believing in capability and the vision of the proposed CEO and rest of the management team who have applied for ICTAI funding. Alternatively, ICTAI Inc. may also search and select a CEO and entrust her with running a particular ICTAI.

The initial funding for a ICTAI would constitute the seed funding (in the range of INR 200 crore – INR 500 crore per ICTAI) for first 5 years to cover the major operational expenses of the ICTAI, in addition to the physical infrastructure and technology / computing infrastructure. The seed funding should be a combination of funding from the government and private participation, preferably in equity/grant sharing mechanism. Corporates should be free to take a predetermined amount of ownership. One corporate should be partnered with for each ICTAI based on a "*challenge method*". The incentive for corporates to participate would arise from the following points that the government should endeavour to ensure:

a) Access to high quality training data

b) Computational and physical infrastructure

c) A chance for staff at corporates to be part of a national mission and work on challenging problems with a higher gestation period than traditional commercial problems

d) Ability to count expenditure incurred towards CSR

e) Boost in visibility received by working on AI for social good directly with top government institutions

The ICTAIs should raise additional funding beyond this base funding, through philanthropic / private contributions, preferably in equity sharing mechanism as well. ICTAI Inc. may also award further contributions to ICTAIs based on measures of success.

Each ICTAI should have a Governance Board, comprising of industry leaders, academic luminaries and global thought leaders. The CEO and / or the industry partner (if one is there) may suggest the Governance Board, which may be vetted and agreed by ICTAI Inc. An industry partner with sizeable contribution may have a reasonable say in selection of the Governance Board. The responsibilities of Governance Board will be akin to a Board of Directors / Board of Governors of any suitably large successful corporation i.e. provide overall guidance and direction, and oversee the output for the ICTAI. While ICTAI Inc. may lay down the success criteria for ICTAIs in general, specific targets for each ICTAI should be decided between the Governance Board and leadership team of individual ICTAIs.

The ICTAIs should ideally be located close to top engineering institutes / close to large cities, so that attracting the best talent from across the world is plausible. ICTAIs should also seek to leverage the availability of students / research scholars from the engineering institutes close to them for specific projects etc. (through internships etc.) to build the talent pool for future. ICTAIs should aim to involve the successful AI researchers / practitioners of Indian origin from across the world in either full time or advisory capacity. If finances are a constraint for ICTAIs to attract the best talent, ICTAI Inc. may consider National AI Fellowships for the benefit of ICTAIs.

PRAIRIE Institute, recently established by the French government in collaboration with academia and industry, is a good potential implementational model for CTAIs.

Box 14: The PRAIRIE Institute

The PRAIRIE Institute (PaRis Artificial Intelligence Research Institute) is a collaboration of industry and academia supported by the French government to create an institution which becomes an international benchmark in AI.

The partners (CNRS, Inria and PSL University, together with Amazon, Criteo, Facebook, Faurecia, Google, Microsoft, NAVER LABS, Nokia Bell Labs, PSA Group, SUEZ and Valeo) in PRAIRIE institute are pursuing three goals:

- to drive progress in fundamental knowledge creation in AI (AI) freely distributed among the international scientific community;
- to take part in solving concrete problems with a major application-related impact;
- to contribute to training in the field of AI.

The five-year objective is to bring together AI scientific and industrial leaders and make the PRAIRIE Institute a world leader in AI.

The aim of the PRAIRIE Institute is to act as a catalyst for exchanges between the academic and industrial worlds, to train new generations of researchers in AI and to play a role in leading and coordinating the community. Transfer and innovation will be among its duties, along with scientific progress. The work done will highlight an integrated approach to the two traditional branches of research:

- upstream research, calling on partner facilities of excellence in France and abroad;
- research focusing on companies and applications, drawing on industrial partners, who are often also world leaders in their fields.

Integration between research topics will facilitate synergy between the two branches of the PRAIRIE Institute and will enable researchers to make the transition easily from one to the other.

At an international level, the PRAIRIE Institute will draw on a network of partnerships with centres of AI excellence to promote exchanges and leverage impact. Collaborative agreements have already been signed with the Center for Data Science at NYU, the AI laboratory of UC Berkeley (BAIR), the Robotics institute at Carnegie-Mellon University in Pittsburgh, MILA in Montréal, the Max Planck institute in Tübingen, the CIIRC (Czech Institute of Informatics, Robotics and Cybernetics) and the Turing Institute in London.

Common Compute Platform for CoEs / CTAIs

For COREs and ICTAIs to discharge their duties effectively, pooled cloud infrastructure for AI applications should be made available. All the COREs and ICTAIs should be connected to the National Knowledge Network (NKN) and from there via a very high-speed link to a pooled cloud computing environment. This approach will reduce infrastructural requirements due to pooling efficiencies and reduce operational and maintenance costs while keeping national data secure (Box 15: AIRAWAT)

Box 15: AIRAWAT (AI Research, Analytics and knoWledge Assimilation plaTform)

Scope

AIRAWAT will be a cloud platform for Big Data Analytics and Assimilation, with a large, power-optimised AI Computing infrastructure using advanced AI processing. The proposed Infrastructure will be equipped with facilities for world's leading machine learning including deep learning, high performance high throughput supercomputing, infrastructure to store, process, simulate and analyse big data sets like images, video, text, sound, speech. AIRAWAT will support advancement of AI-based developments in image recognition, speech recognition, natural language processing for research, development and creation of varieties of new applications for the support of advancements in the fields of Agriculture & Healthcare.

Infrastructure Requirements

1. AI Infrastructure: High throughput processing supercomputing systems to train the machines learn the data sets using AI deep learning. We propose to set up a 100-AI petaflop computing Infrastructure, like the one as JAPAN's ABCI AI Supercomputing facility.
2. Multi-tenant multi-user computing support through resource partitioning and provisioning, dynamic computing environment deployment, etc.
3. Energy-saving, high Teraflops per Watt per U rack space designed computing systems with low TCO
4. Deep Learning Software stack – Training and Inferencing development kit, frameworks, libraries, cloud management software and portal for data labelling, data analytics, data transfer & data model exchanges
5. Low latency High bandwidth network
6. Mass Storage System to collect, store and share multi-petabytes of big data

Utilisation

Big data Labelling, Annotating, Anonymisation, Analytics, Skills development, job creation, support R&D, Academia, startups, entrepreneurs and end users for agriculture & healthcare advancements. intelligence.

***Credit**: nVIDIA*

Pursuing excellence in research in Artificial Intelligence

To achieve technology leadership in AI, India also needs to pursue "moonshot" projects – ambitious explorations that aim to push the technology frontier and that would require the pursuit of world class technology development and leadership in applying AI technologies to solve some of the biggest challenges. A potential project could be tested on a twin criteria of whether it is (a) a new technology or scientific area that has emerged or gained traction and has the capability of solving, often in new ways, practical problems of importance, and (b) whether it addresses emerging user needs that existing and available technology solutions cannot address. An example could be development of advanced natural language processing (NLP) infrastructure for Indian languages, with features of sentiment and semantic analysis, or imaging biobank for early cancer detection.

Some experts believe that such ambitious projects should be driven by the industry, since they have the requisite resources, technical expertise and are attuned to practical use cases; and should rarely be pushed forward by government initiatives. Google's advances in NLP, as underlined by Duplex, is one such example of private sector led innovation that pushes the realisable boundaries of technology. However, the inherent risk and ambition of such projects does not necessarily lay themselves well to the functioning and methods of most corporations.

ICTAI Inc., the overall governing body for ICTAIs, could be the DARPA (see Box 16) equivalent of AI research in India. ICTAI Inc. should have the full autonomy on selecting the projects of importance and executing these projects, and should be trusted with a separate budget for pursuing such projects. Some of the key elements of execution include:

a) *Projects with pre-defined durations and tenures*: Although ICTAIs could adopt the moonshot approach to select projects, the projects shall have pre-defined durations for efficient management, timely realisation of research deliverables and financial sustainability. The type of project could govern the timeline, for example a project that requires integration of existing AI technologies to solve a specific problem could have a shorter duration, whereas projects which would require a unique technological development could be allocated a timeline suitably.
b) *Teams of contractors*: Specialised teams of contractors who have proven credentials in advanced technological development shall be chosen to undertake the projects. Focus could be laid more on the qualitative impact of past research activity, and less on the size or global recognition.
c) *Special emphasis on selecting the right project leaders*: Project leaders with the right skill sets to lead projects of national importance shall be key to drive the research output of projects undertaken at ICTAIs. They shall be people with proven capabilities in applied research, and the acumen to deliver technology in the form most suitable for India and the developing world.
d) *Independence*: Independence to select projects, design timelines and deliverables and allocate budgets shall fall within the ambit of the ICTAIs with no intervention of any government agency. Selection of the project teams based on credentials while collaborating with premier research institutions with full independence shall ensure a robust technological development that creates impact.
e) *Fail fast model*: ICTAIs shall be given the autonomy to shut down projects and shift resources to other operational projects. With the moonshot approach for project selection, it is plausible that certain research projects do not provide expected results in terms of efficiency, scalability or cost. In such cases, diverting resources to other operational projects with higher probability of research output delivery would ensure financial efficacy.

Projects could be housed at one ICTAI, but could involve working across multiple ICTAIs and COREs.

The algorithms and data used in an AI powered application are key elements in ensuring operational success. Therefore, it is imperative that the Intellectual Property Regime in the context of AI be robust and enforceable for innovators to have the confidence that they will be able to make profits from and take credit for their work. This is essential for the promotion of innovation, entrepreneurship and core and applied research in the field of AI.

However, IP in the context of AI has some key differences with IP in the context of generic computer programs or other content due to the way such algorithms are designed and trained with large datasets. The data fed to an AI algorithm during the training step is key to its success. Keeping this and other issues in mind, the government may set up a task force, comprising jointly of Ministry of Corporate Affairs and DIPP to examine and issue appropriate modifications to the IP regulatory regime pertaining to AI.

Box 16: Defense Advanced Research Projects Agency

Pentagon's Defense Advanced Research Projects Agency (DARPA) was created in 1958 as response to increased US concerns that Soviet Union may have achieved technological superiority following its successful launch of Spuntnik. DARPA's founding mission was simple and nimble, "to prevent and create strategic surprise".

For the past 60 years, DARPA has been the beacon of scientific breakthroughs and radical innovations. Amongst its success stories include the internet, RISC computing; global positioning satellites; stealth technology; unmanned aerial vehicles, or "drones. While originally created for US military, the agency has played a pivotal role in giving genesis to several multi-billion dollar industries.

What is even more impressive is DARPA's agility and swiftness despite a small team and modest budgetary allocations. DARPA's total support staff is only 120 strong, and its annual budget is only USD3 billion. DARPA accomplishes most of its programs through a "specialised project" model – assembling project specific teams of experts from universities, industry, government and non-profit, with well-defined targets and tenures.

The three essential elements of DARPA are:

a) *ambitious goals:* the projects pursued by DARPA are designed to solve real-world problems or create new opportunities. The problems are such that they can't be solved without pursuing new frontiers or catalysing new developments. Urgent need for application underlines the problems.

b) *temporary project teams:* teams of world-class experts from industry and academia are assembled to work on well-defined projects of relatively short duration, led by accomplished technical managers. Given the intensity, sharp focus and reputation attached, these projects attract high caliber talent who achieve extraordinary levels of collaboration.

c) *independence:* DARPA has complete autonomy in selecting and running projects.

Source: Harvard Business Review article "Special Forces" Innovation: How DARPA Attacks Problems published October 2013

CERN for AI

Gary Marcus, Professor of Psychology at NYU, mooted the idea of a CERN for AI at the #AIforGood Summit in Geneva in June 2017. In an OpEd for the New York Times, Gary further elaborated on the idea:

"I look with envy at my peers in high-energy physics, and in particular at CERN, the European Organisation for Nuclear Research, a huge, international collaboration, with thousands of scientists and billions of dollars of funding. They pursue ambitious, tightly defined projects (like using the Large Hadron Collider to discover the Higgs boson) and share their results with the world, rather than restricting them to a single country or corporation."

Box 17: European Organisation for Nuclear Research aka CERN

CERN, established in 1954 by 12 founding European nations, had its origins in establishing a truly international scientific collaboration to pursue world class research in particle physics. Among the biggest of the several achievements to come out of this remarkable collaboration, home to the world's biggest atom smasher (the Large Hadron Collider) include the development of World Wide Web and the discovery of the Higgs boson (the 'God Particle'), W and Z bosons, Antimatter etc.

CERN's current 21 member states each contribute to the overall CERN budget (CHF1,240 million in 2013). Special contributions are made to specific projects, like Large Hardon Collider, by interested Host States and non-member states. The biggest chunk of CERN's budget is spent on the construction of its enormous scientific installations, most of which are too expensive for a single country to afford.

The current state of AI, which is focused on solutions to narrow applications, and grappling with questions around next phases of development, ethics, security and privacy may benefit from a global public AI research institution advancing the field for the good of humanity, the "People's AI"[30].

To achieve #AIforAll, which gives the mandate for inclusive AI for the world, the Government of India should take the lead in bringing together the relevant parties to create People's AI, the CERN for AI. India has a proven track record for leading projects with ambitions of greater good. India is already playing a leading role in climate leadership, with Hon'ble Prime Minister Narendra Modi vowing to go "above and beyond" India's commitment on Paris Agreement on climate change. Similarly, India has been a pioneer in a sustained push for clean energy revolution by leading the International Solar Alliance, and setting an ambitious target of 100GW of installed solar energy capacity by 2022. With 20GW of installed solar capacity, India is well and truly on its way to achieving this target. With the Government of India's focus on inclusive growth which saw several transformational reforms in the last few years, India is poised and most suited to wear the mantle of leading the #AIforAll movement. Furthermore, India offers the best possible test bed and a plethora of use cases for building AI solutions fulfilling the inclusive AI criteria, be it in healthcare, education or agriculture.

While the modalities of funding and mandate for this #AIforAll should be the subject of further deliberations, the proposed centre should ideally be funded by a mix of government funding and contributions from large companies pursuing AI (GAFAM, BATX etc.). Where should this centre be located? Nowhere and everywhere. While the CERN had the requirements of physical facilities such as the Large Hadron Collider, #AIforAll could be distributed across different regions and countries. The Government of India, through NITI Aayog, can be the coordinating agency for initial funding and setting up the requisite mandates and human and computing resources.

The mandates of the #AIforAll should include challenges of AI that are common to other countries and large corporations and startups alike, and the foundational components that could truly make AI inclusive and good for all. Suggestive technology topics could include:

a) General AI;

b) Opening the Black Box / Explainable AI;

c) Advanced anonymisation protocols for data security and privacy;

[30] *Medium.com "The People's AI: democratizing AI research and development"*

d) Ethics in AI; and

e) AI approach to solving world's biggest problems in healthcare, education, urbanisation, agriculture etc.

A start has been made by OpenAI, set up by the likes of Elon Musk and Sam Altman, with a mission to discover and enact the path to safe artificial general intelligence. Similarly, Declaration on Cooperation in Artificial Intelligence was signed by 25 states of European Union on 10th April 2018. What is needed though is a commitment and collaboration of truly international standards to develop and ensure #AIforAll.

Skilling for the AI age

Getting India ready for the AI wave

History suggests how technology has disrupted the nature of jobs and the skills required to perform them, requiring the global workforce to continuously adapt. Advent of AI has accelerated this disruption to a pace that has not previously been seen, due to the wide range of capabilities it offers and speed at which it is developing.

NASCCOM predicts that by 2022, a startling 46% of the Indian workforce will be engaged in entirely new jobs that do not exist today or jobs that have radically changed skill sets[31]. Some other sources estimate that demand for AI and machine learning specialists in India is expected to see a 60% rise by 2018[32]. In the data domain as well, an independent study estimated that India will face a demand-supply gap of 2,00,000 data analytics professionals by 2020.

In the IT-BPM sector, traditional software developer roles are set to transition to roles such as computer vision engineers, Robotic Process Automation (RPA) engineers and cloud architects, among others. At the same time, completely new job roles such as language processing specialists and 3D modelling engineers are set to arise as the technologies are increasingly adopted and deployed[33]. With AI, such transition would move beyond the IT sector and affect sectors such as education, health, agriculture, finance etc, requiring the underlying skill sets.

[31] *Future of Jobs in India: A 2022 Perspective, 2017*

[32] *KellyOCG, India*

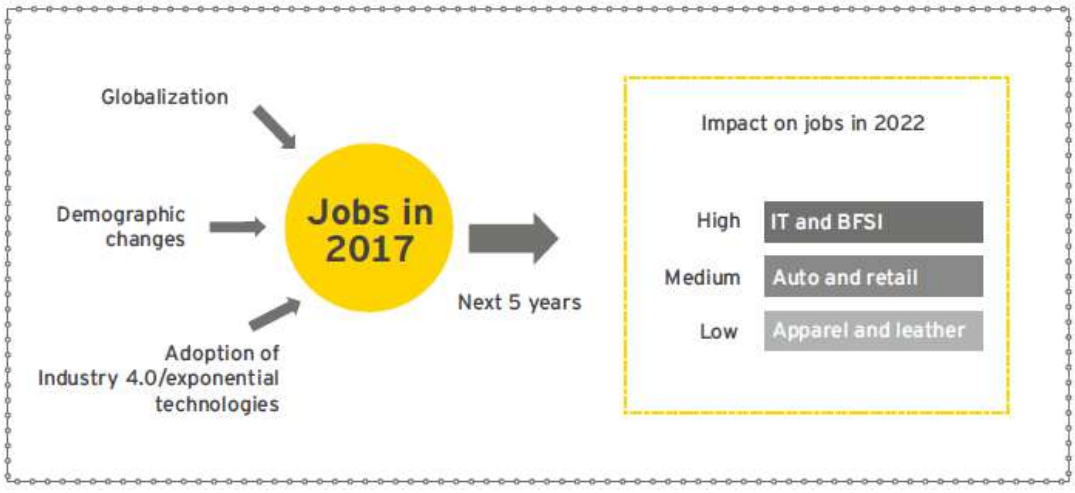
[33] *Future of Jobs in India: A 2022 Perspective, 2017*

Figure 21: *Transition in jobs and skill sets*

Source: Reworking the Revolution, Accenture

The demand for new-age jobs is accelerating in India, and can be attributed to three major factors[34], viz. increased adoption of technology; shift in market demographics and de-acceleration of globalisation.

Figure 22: Impact of primary forces on jobs in 2022



Source: Future of Jobs in India: A 2022 Perspective, 2017

As technology adoption increases across sectors, so will the demand for skills required to implement them. India's growing middle class and millennial demographic form a major part of the market for technology enabled products. As the size of these demographics increases, so will the demand for these products, and thus the workforce that can enable their creation. Increased globalisation played a major part in the creation of jobs in the services sector. However, as the permeation of automation increases,

[34] *Future of Jobs in India: A 2022 Perspective, 2017*

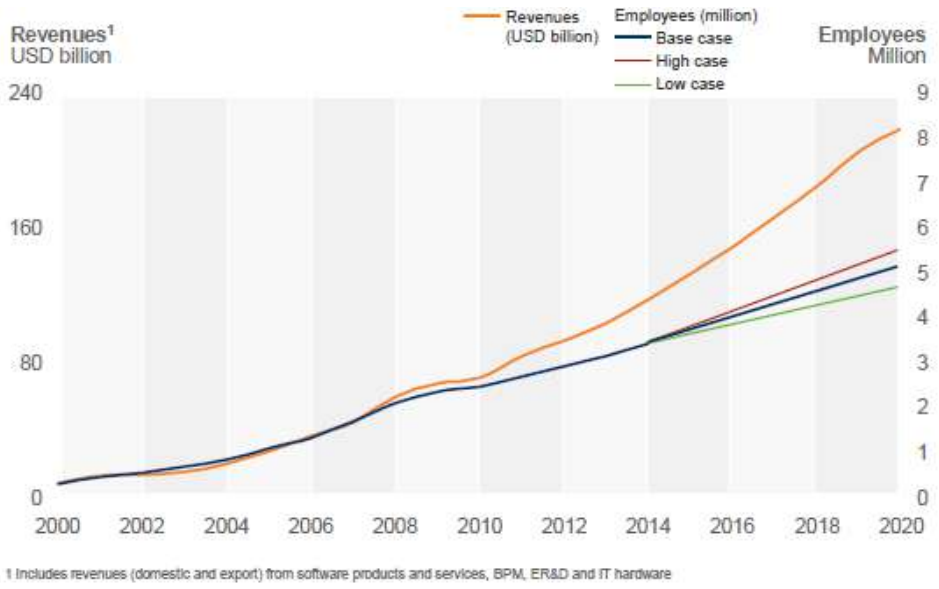
and protectionism among previous markets for these services becomes predominant, it is important to recognise that these jobs will see a large overhaul.

India may appear to be relatively well positioned to take advantage of the AI disruption by virtue of its advanced IT sector and large youth demographic potential to establish itself as the future hub for AI related activities. However, given the poor availability of qualified faculty and researchers, this advantage could fast transform into a liability without urgent government interventions towards promoting access to such skills. This is a critical component of AI development, and should be a national priority.

Changing nature of the global service sector, uncertain impact of automation in manufacturing, and poor infrastructure can pose a challenge to the task of enabling up-skilling or re-skilling.

Shifts in the outsourcing needs of developed economies have the potential to impact India dramatically. As multinational companies move toward 'in-shoring' or 'no-shoring' due to development of automation technology and increasingly protectionist measures globally, it is of critical importance that the outsourcing industry (which currently accounts for roughly 15% of India's total labour force) adapts to these changing needs accordingly.

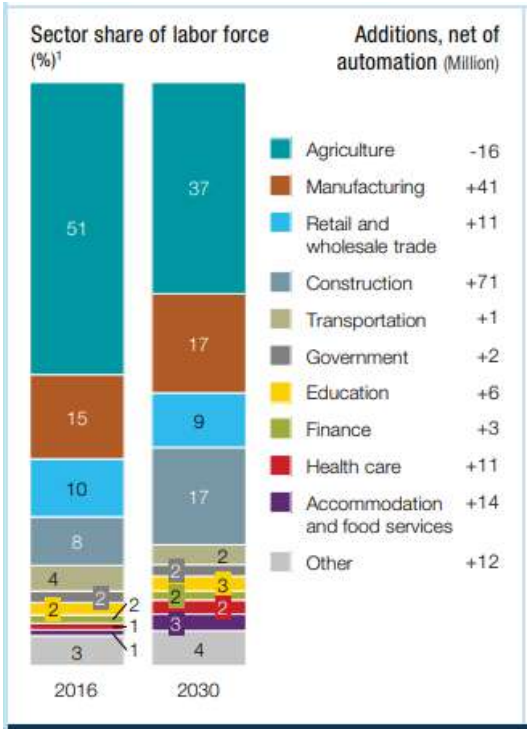
Figure 23: *India IT-BPM sector*



Source: NASSCOM, ICRIER, McKinsey analysis

As indicated below, automation could affect even traditional sectors in a variety of ways. Agriculture is predicted to see a net decrease in jobs due to automation, while the job opportunities in the construction sector are predicted to improve significantly. These large scale shifts in employment depend largely on the nature of technology being developed as well as deployed, and are very hard to predict for any informed policy intervention.

Figure 24: *Impact of automation*



Source: McKinsey, Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation

Multiple studies have underlined the lack of employment readiness of STEM graduates, thus highlighting the poor standard of education in engineering colleges in India. As per some estimates, almost 80% of engineering graduates are unemployable on graduation. Low availability of specialised faculty; lack of flexibility in curriculum revisions in engineering and computer science courses to keep up with rapid advancements in technology; and low levels of interdisciplinary research in AI related fields to facilitate an AI education for non-computer science engineering (and vice versa), are just some of the factors that have led to this scenario.

At the school level as well, poor outcomes in Maths and reading are particularly troubling, since these subjects form the foundation of knowledge required to move to an AI related education, and later jobs in the domain.

Recommendations

For addressing issues relating to skilling, a two-pronged approach is warranted, one set of interventions aimed at the workforce and the second for the students.

1. Workforce

Re-skilling of the current workforce will require integration with relevant existing skilling initiatives, building of new platforms that can enable improved learning, and novel methods of allowing large scale employment generation through promotion of AI.

a) *Incentivising creation of jobs that could constitute the new service industry:* To tackle the challenge of shifts in the services industry, it is important to identify and promote creation of jobs that may replace traditional IT-BPM sector jobs in the future. These jobs would ideally be a part of the AI solution development value chain, but require a relatively low level of expertise so as to create employment at scale. Tasks such as data annotation, for example, have the potential to employ a large quantum of human resources, and serve countries all over the world in otherwise capital

intensive projects. Tasks such as image classification or speech transcription require low levels of expertise and present an opportunity to exploit labour cost arbitrage to serve companies globally. Specific policy interventions could be considered like tax holidays, inclusion under CSR activities, etc. to help solve the dual problem of workforce job displacement and creation of expertise in fundamental sections of the solution development value chain.

Box 18: Impact sourcing through Samasource

Samasource is a pioneer in impact sourcing of technology services, the practice of hiring people from the bottom of the income pyramid and directly raising them out of poverty by providing digital work for companies like Google, Walmart, eBay, and many startups.

About two-thirds of their work is in managed services for image capture and annotation. They need a large pool of skilled, detail-oriented workers to identify the appropriate images or video frames and manually tag the scene, specific regions (a forest), or objects (a wolf) with the specific keywords the customer requires for their application.

Samasource quotes its top objective as 'creating technology jobs for the impoverished while delivering the highest quality work in a cost-effective manner'. This guided Samasource to focus on East Africa, where the majority of the population has completed secondary or high school, and has a good understanding of basic English language. They have a dedicated training organisation that works within the local communities to identify qualified men, women and youth in need and assess them for key skills. For those with strong visual acuity, there is general computer and business skills training and a 3-week dedicated machine learning and image annotation training track with example projects and ongoing assessments that lead to graduation

As workers get moved to projects, there is another set of qualification tests and project-specific training. Once they pass those, they are moved into production for the project and their work is reviewed daily by the QA team to correct mistakes and resolve interpretation or requirements issues (do you want to tag a traffic light facing away from the camera?). These reviews are done both locally and in the US, allowing real-time feedback with many checkpoints throughout the project lifecycle, reinforcing Samasource's tightly managed approach to quality.

Samasource reportedly provide jobs and digital skills training to people below the poverty line in Kenya, Uganda, Haiti, and India. Since the company started in 2008, Samasource has been able to hire nearly 8,000 people. As a result, they've could transition themselves and their dependents out of poverty and provide living wages, transforming over 30,000 lives.

b) *Recognition and standardisation of informal training institutions:* The increasing demand for AI or data related job positions has not gone unnoticed by the Indian workforce, with a large percentage of them opting for training institutions to bridge their knowledge gaps. In technology hubs such as Bengaluru, this has led to many traditional IT training institutions establishing courses in new age technologies. However, their standard of education is hard to assess for companies looking to hire. Implementation of recognised certificate courses through higher education institutions could be a major boost to recognising resources spent on re-skilling, and holding these institutions to standards in delivery of knowledge. International School of Engineering (INSOFE), for example, provides certification recognised by the Language Technologies Institute of Carnegie Mellon University (CMU) for a post-graduate program in data analytics and optimisation. Jigsaw Academy's data science post-graduate program gets its students certified from the University of Chicago Graham School.

Integration and application of existing standards such as laid out by the National Skill Qualification Framework (NSQF) should also be explored. Given that standards in areas such as big data exist but are not used by institutions for certification also highlights the need for them to be designed in closer collaboration with the private sector.

c) *Creation of open platforms for learning:* Initiatives such as the NASSCOM Future Skills Platform will play an instrumental role in large scale dissemination of requisite skills to some major sections of the employed workforce. Online and self-learning platforms, such as Coursera and edX, are able to connect learners to the best universities and institutions from around the world, and can play a crucial role in this scenario. There is need to bring out guidelines for promoting these while ensuring uniformity, standards and usability. As in the promotion of MOOCs, large scale deployment and adoption of these platforms requires stringent measuring of quality, and recognition of their certification.

d) *Creating financial incentives for reskilling of employees:* Initiatives in reskilling of employees or allowing employees to undergo reskilling initiatives have a high opportunity cost for private companies, and may affect their willingness to let their employees engage in the process at scale. It is thus suggested that co-funded models between the government and companies be explored, in the IT sector particularly. Financial incentives for private companies could include payroll taxes which are dedicated to subsidising training opportunities, income tax deductions for companies participating in reskilling initiatives, special taxes to be paid if a minimum training budget is not disbursed, as well as public grants for subsidising training especially for smaller sized firms. Considering also the time required for reskilling, and the cost it entails to employers, financial incentives may also be tied to mandatory allocation of time for reskilling activities by companies for their workforce. However, in the absence of standardisation of training modules and institutions, such initiatives could be prone to misuse.

2. Students

The education sector needs to be re-aligned in order to effectively harness the potential of AI in a sustainable manner. In primary and secondary schools, there is a need for transition to skill based education in subjects relevant to AI. Often criticised for being overly knowledge intensive, Indian education is in urgent need of transition particularly in subjects relevant to STEM, or computer based education. As jobs based on technology become prominent, so will the need to develop applied skills in a continuously changing environment.

Increased amount of project related work across education levels, promoting schemes like the establishment of ATLs (Atal Tinkering Labs) in schools, necessary change in curricula in schools, are some of the steps that need to be considered to promote early adoption of technology organically.

In higher education institutions there is need for increased collaboration between industry and academia through creation of channels of communication between faculty and industry to promote exchange of ideas and expertise. Various avenues of collaboration need to be explored, including workshops, incentives for guest lectures by professionals and institutional arrangements for regular re-design of courses in collaboration with the private sector.

Lack of qualified faculty that poses a serious problem in the present scenario can be addressed through innovative initiatives like *credit-bearing MOOCs* (Massive Open Online Courses). Acceptability and adoption of these decentralised teaching mechanisms can be ensured through prescribed certification in collaboration with the private sector and educational institutions. Initiatives such as the *SWAYAM*

platform[35], are in the right direction but these need to be further reinforced through additional investment and collaboration with the private sector and educational institutions in order to meet the market demand.

Countries like the USA, that have a thriving student and research community engaged in AI, have effective bridge courses, generally at post graduate levels for non-related fields. Thus at a post-graduate level, the focus shifts to creating *incentives to conduct research in domestic institution*s and allow *cross-disciplinary collaboration and education*. Similar initiatives are required in Indian educational space with bridge courses in AI for post-graduates in non-computer science or data science domains to encourage cross-domain research and the 'AI + X' paradigm. One year courses could be explored that will enable students of a range of subjects to build foundations of knowledge in the AI space, and effectively apply domain expertise to solve pressing problems.

This is an ever evolving area and needs to go beyond the one time solutions with an institutional framework in place that is responsive to changing scenarios. Such a framework would enable a more sustainable and customised solution. A *standing committee or task force* comprising of all stakeholders may be constituted by the government with the objective of examining and reporting the changes in employment caused by AI in India. This taskforce would consider not only IT related activities, but the job ecosystem as a whole. As AI continues to evolve, this will facilitate evidence based decisions in funding of educational institutions, promotion of specific sectors, and channelling of human resources to be most efficiently utilised.

[35] *developed by the All India Council for Technical Education (AICTE) and the Ministry for Human Resource Development (MHRD)*

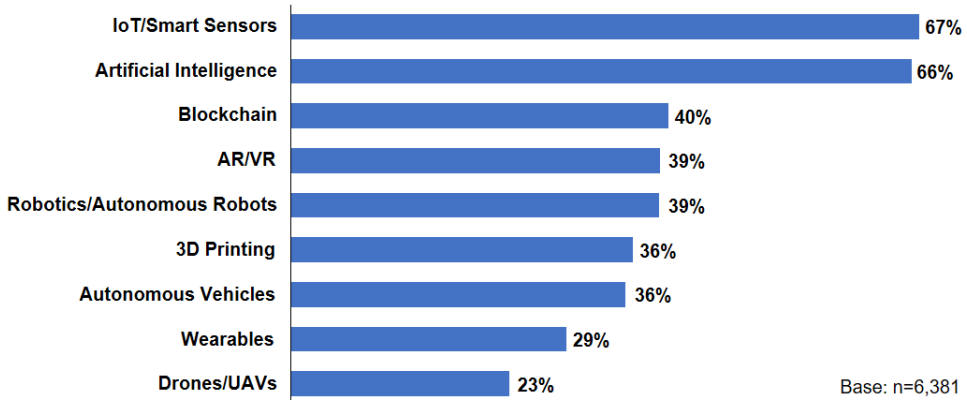
Accelerating Adoption

AI across the value chain

Global Context

Adoption of AI globally is still in its nascent stages, but growing rapidly. A 2017 survey by Statista finds that 78% of firms globally are either using AI extensively, or have plans for use in near future. Firms in China and the US especially, are proactively engaging with their research communities to enable early adoption and position themselves competitively.

Figure 25: *Survey of global firms on adoption of emerging technologies in 2017 (In which of the following new and emerging technologies is your organisation making investments over the next year?)*



Source: Accenture

A McKinsey report on "AI: The next digital frontier" concludes that AI adoption outside of the tech sector is at an early and often experimental stage and deployment at scale has been rare. In the survey of 3,000 AI-aware top management executives, across 10 countries and 14 sectors, only 20% use any AI related technology at scale or in a core part of their businesses.

As the field is rapidly evolving, investments made in core AI research and product development is another indicator of adoption. The McKinsey report estimates that organisations invested around USD26 – USD39 billion in AI while USD6 – USD9 billion was invested in startups in the year 2016.

Adoption of AI in India has been slow and remains limited. Estimates indicate that only 22% of the firms in India use AI in any business process[36]. Indian startups have been able to raise just USD87 million in 2017, as against over USD28 billion raised by the Chinese startups in 2017[37].

Low adoption of AI technologies in India is particularly troubling, given the country's prominence in the global IT industry that could have given it the natural first mover's advantage in AI. However, the IT industry in India has remained content in delivering traditional IT services and has been slow to adapt to new digital technologies compared to its counterparts in China and the US.

[36] *Intel and IDC Survey Report, dated*

[37] *XinhuaNet "AI sector sees big investment, financing in 2017"*

Despite its sectoral leadership and programming talent, India's IT industry with over USD160 billion annual revenue, has yet to build pioneering AI / ML capabilities commensurate with its potential. Nevertheless, the top 5 IT service companies have begun to use AI to cut costs and automate business processes. Wipro has built *Holmes*, an AI platform that deploys "bots" to carry out repetitive and mundane tasks. TCS has created its own AI platform, *Ignio*, and Infosys has built *Nia*, improving upon its earlier *Mano* platform. While promising, these early efforts remain far from revolutionary.

The limited success of Indian technology players to effectively adapt and carry forward the AI revolution suggests the need for government intervention to promote AI adoption, lest India lose the chance to secure a prominent position on the global AI map. While acknowledging the need to promote AI, governments at different levels, along with their various instrumentalities, should adopt proactive measures to accelerate AI adoption in various processes.

The major market segments for the increased AI adoption are:

(a) *Private enterprises*: mostly driven by market and enterprise considerations,

(b) *Public Sector Undertakings*: imperative to drive up the operational efficiency of PSUs, and

(c) *Government*: improve process efficiency, reduce human discretion, eliminate middlemen, advance prediction, pro-active and predictive service delivery to citizens.

The steps taken by the government, in terms of incentives or in providing infrastructural platforms, will have varying control and influence on all these segments – least on private enterprises to most on government agencies.

AI adoption in India will face the following challenges:

(d) *Lack of adequate talent to build and deploy AI systems at scale*. An estimate claims that only 4% of AI professionals in India have worked on emerging technologies such as deep learning and neural networks. There is also a significant gap of PhD research scholars in the field.

(e) *Difficulty in access to industry specific data required to build customised platforms and solutions currently concentrated in the hands of a few major players*. It is difficult for new entrants to deliver tailor made services that can compete with data rich incumbents such as Facebook or Google. This phenomenon results in the creation of a virtuous cycle which reinforces the hegemony of the big few, creating a huge entry barrier for startups.

(f) *High cost and low availability of computing infrastructure required for development, training and deployment of AI based services*. Cloud infrastructure, though growing rapidly, has limited capability. Also lacking are AI-as-a-service models of cloud platforms. Lack of infrastructure has led to many Indian AI startups to incorporate their business outside the country, which makes AI outside the reach of Indian researchers in government labs and many industries. Initiatives like GI Cloud (MeghRaj), are in the right direction.

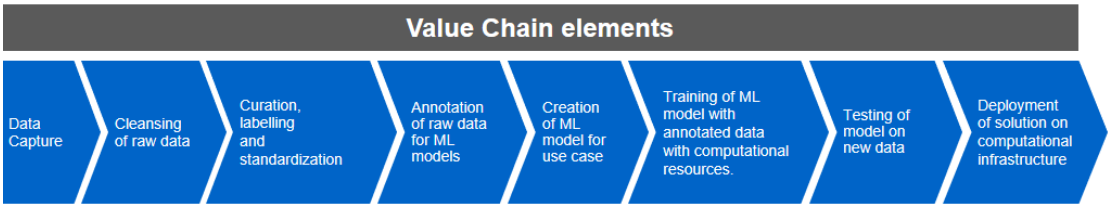
(g) *Low awareness of AI* for resolving business problems in most public enterprises and government agencies, especially given the scarcity of AI professionals, is obstructing adoption.

Following specific initiatives are recommended for promoting AI adoption in the country:

Recommendation 1: Creating a multi-stakeholder Marketplace

To encourage the development of sustainable AI solutions at an appropriate price point for sectors such as health, education, and agriculture, it is necessary that a level playing field be ensured and a supportive environment be created for all players in the value chain. The development of any working AI-based product is a long process with very different specialised activities that are necessary for final delivery, just like any other product or service value chain.

Figure 26: *AI Value Chain*



It is exceedingly difficult for a small or medium business / startup to vertically integrate all these processes and have the internal capacity to deploy them simultaneously before even entering the market. Therefore, this acts as a barrier to entry for new players. On the other hand, for different firms to take up different activities and still deliver useful products, a strong, stable price discovery mechanism has to exist for incentives to align and sustainable business models to come up for the different parts of the value chain. Another well documented and substantial barrier to entry is the difficulty of acquiring the raw material (data) in the first place since as noted above, most usable data is in the hands of a few players.

Towards providing a level playing field, addressing information asymmetry, and incentivising and simplifying collaboration between the various stakeholders in the AI ecosystem, a marketplace may prove to be a potent tool provided it enables the following:

(a) *Discovery of required AI component*: ability to reference assets, be it data or ML models, and services, such as annotation, and enable curation and rating of these assets.

(b) *Platform to execute transactions*: with mechanisms to exchange value of a specific good and service for currency or subscription models for APIs.

(c) *Verification of transactions*: ability to verify the occurrence of transactions and the receipt of goods or services in exchange for currency.

Such a marketplace model would benefit by:

a) *Reducing asymmetry of information:* With a reasonable volume of transactions, aberrations caused due to data oligopoly will reduce, thus reducing the spread in pricing and effectively incentivising both data owners and AI model creators.

b) *Encouraging specialisation in different niches by firms and creation of novel business models:* Easy availability of raw components required at different stages of the AI solution development cycle would allow firms to focus on specific problems, rather than attempt to build capability across the value chain. For example, a new segment of startups could entirely focus on the problem of curating raw data to be put on the marketplace, rather than attempting to do all the above at once.

Consequently, such startups will see a better market for VC funding as they will be able to substantially mitigate their business risks, if an easily accessible, formal market for raw data exists.

c) *Unlocking new sources of data and enabling more efficient use of computational and human resources:* Estimates suggest that only 1% of data today is analysed due to its existence in various unconnected siloes. For example, medical imaging / diagnostic centres today are collecting valuable data. It is currently not used to its potential as there is no way for diagnostic centres to build predictive disease models without hiring AI experts, renting computational expenditure, and indeed shifting gears completely to become an AI company. In the presence of a formal market with a fair mechanism of price discovery, such diagnostic centres would have an incentive to collect this data, have it curated, and place this data in the market with appropriate permissions and safeguards. Concurrently, machine learning experts in AI firms would be more productive as they could focus on the problem of creating the model rather than that of sourcing the data and curating it.

d) *Address ethical concerns regarding data sharing:* Today, transactions and buying of data occur in an informal marketplace. However, there is no mechanism currently to ensure that appropriate permissions are taken from the actual data owners before data custodians share the data. Going back to the example of diagnostic centres (which are data custodians in this case), it would become necessary for firms to take consent of individuals getting imaged to aggregate and sell their data. This could create mechanisms through which individuals are ultimately compensated. In this way, the informal market for data could be nudged towards entering the formal economy.

Recommendation 2: Facilitating creation of large foundational annotated data sets

In India, lack of annotated data in the domestic context has emerged as a major impediment in development of AI solutions for both startups and core research alike. Availability of general data corpora which can be applied across product functions can serve to provide a ready source of data (in 'plugin' mode) to startups and enable solutions customised to the Indian context. There is evidence to suggest that ready availability of large corpora can *spur research and innovation* in the field of machine learning.

Figure 27: *Breakthroughs in AI*

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough:		**3 years**	**18 years**

The average elapsed time between key algorithm proposals and corresponding advances was about 18 years, whereas the average elapsed time between key dataset availabilities and corresponding advances was less than 3 years, or about 6 times faster.

By its very nature, this task is laborious and despite innovations in automatic annotation, such as the development of *GoogleNet object classification,* human annotation and training of data sets are not

expected to be replaced in the near future. This is best exemplified by the thriving investments and startups in this area. As lately as 2017, Alegion, Scale, CloudFactory, Mighty AI, and CrowdFlower, all companies that enable human intervention to data annotation, have received about USD50 million in investment funding.

Given the nature of this task, it is necessary for the government to explore assistance in building of large corpora of data across domains, as a means of laying the foundation for startups and enterprises to build applications and services which are tailor-made to the Indian context, and in the process lowering entry barriers for startups and academia while also encouraging international expertise to focus on problems in the Indian context.

In areas such as native language NLP (Natural Language Processing) for diverse Indian languages, for example, the funding for creation of these data sets can add incremental value to existing services across many domains, ranging from e-commerce to agricultural advisory. Co-funding by the government would also enable enforcing of standards across development of data sets, and allow interoperability at a large scale. A collaborative approach where the government acts as a catalyst is recommended as a way forward.

Recommendation 3: Partnerships and collaboration

AI is a highly collaborative domain, and any framework aimed at promoting AI needs to be aligned accordingly. A multi-pronged approach, involving various stakeholders and promoting a collaborative approach is required for promoting development of AI tools as well as adoption of AI in different fields of activity.

Perkmann and Walsh (2007) studied different degrees of partnership in research and categorised them in three categories, each having different characteristics.

Figure 28: Categories of partnerships in Research

High (Relationships)	Research partnerships	Inter-organizational arrangements for pursuing collaborative R&D, including research consortia and joint projects.
	Research services	Research-related activities commissioned to universities by industrial clients, including contract research, consulting, quality control, testing, certification, and prototype development.
	Shared infrastructure	Use of university labs and equipment by firms, business incubators, and technology parks located within universities.
Medium (Mobility)	Academic entrepreneurship	Development and commercial exploitation of technologies pursued by academic inventors through a company they (partly) own (spin-off companies).
	Human resource training and transfer	Training of industry employees, internship programs, postgraduate training in industry, secondments to industry of university faculty and research staff, adjunct faculty of industry participants.
Low (Transfer)	Commercialization of intellectual property	Transfer of university-generated IP (such as patents) to firms (e.g., via licensing).
	Scientific publications	Use of codified scientific knowledge within industry.
	Informal interaction	Formation of social relationships (e.g., conferences, meetings, social networks).

In India, Academic Entrepreneurship is being taken up through mechanisms such as the development of dedicated e-cells in institutions such as IIT Bombay, IIT Madras and IIM Ahmedabad, among others. Low degrees of partnership, such as direct transfer of knowledge, is being taken up by MNCs and startups that make use of scientific publications and informal interaction to develop. Research partnerships are

also being explored, but in a very limited capacity. Other partnership mechanisms, however, are all but absent.

(h) *Collaboration between Research Organisations* is required to promote the 'AI+X' paradigm where the AI researcher works in close collaboration with the researcher in other domains like healthcare, manufacturing, agriculture, etc. Several successful US universities have co-located departments which enabled close collaboration. DARPA in the USA has close linkages with research facilities housed in academic institutions, and has taken up major initiatives for promoting research and innovation in collaboration with private enterprises and startups alike. In India these domains are not naturally connected, primarily due to the structural nature of both academia and research labs. To promote AI in the country, and fully realise the 'AI+X' paradigm, it is also necessary to enable such cross-sectoral collaboration, e.g. between a medical doctor and a computer scientist, to lead research for application of latest technology in solving medical problems and also making real data available in the process.

(i) *Industry-Research Collaboration* is required to help continuously scale and improve the initial research output based on user feedback from the market. For example, in the domain of search engines, several search datasets like TREC, SMART, etc. were made available by the US universities during the 1970-80s to develop the technology of information retrieval and search. This in turn led to development of several search engines such as Yahoo! and Alta Vista during this period. Over time, this led to improved smarter systems until Google became the market leader in the late 1990s.

(j) *Collaboration with trade bodies and Venture Capitals* is essential for successful functioning of a profitable business which involves various other elements, like deriving the optimal business model, managing human resources, advertisement / marketing and many varied functions depending upon the type of technology and underlying business. Trade bodies and associations are very important for collaborating with other people in similar businesses, as they can share information about the common problems, possible solutions and access to learning from the ground. Such bodies help in collectively identifying new international markets for such products / technologies and in lobbying/negotiating for favourable trade norms with the national and international bodies.

Recommendation 4: Spreading awareness on the advantages AI offers

Another major hurdle in adoption is low visibility for the work being done across the country. Unless known through networks, work is often duplicated without knowledge of previous work done in the area and consultation with experts. There is a need for an AI Database on an online portal for registered people to access and find this information. This database, primarily managed by the government, could serve as a single source of truth for experts and projects being implemented. The resources can include details like researchers stating their expertise, and professionals with hands-on experience on developing AI solutions showcasing their credentials. This knowledge base can also become a forum for sharing various discussion related to research collaboration, and finding relevant collaborators and finding professionals to deliver on AI projects.

There is need to make senior level officials among government agencies, public sector undertakings, and other domestic firms aware of the numerous advantages that AI offers, by organising workshops and live demonstrations of possible AI applications and how its implementation can help augment the human workforce, rather than displace it. The constitution of an annual 'AI Readiness Index' may also be considered, as highlighted in the DIPP Task Force report on Artificial Intelligence. This may also be further expanded to highlight best practices across states, to be shared in various fora. NITI Aayog, given its mandate of competitive and cooperative federalism, could undertake this task. It has already collaborated

with organisations such as Google to organise capacity building training and workshops for government functionaries.

Recommendation 5: Supporting startups

Startups and smaller firms are the engine for growth in a dynamic evolving economy like India, and are constrained in the AI space, thus requiring targeted government interventions.

(a) *Incubation hubs specifically for AI startups* in collaboration with State Governments and private sector stakeholders need to be set up to provide space, and other infrastructure facilities for new startups to incubate along with interacting with other startups at various levels of maturity in order to interact and provide advice.

(b) *Establishment of fund to provide grant funding to startups* to facilitate their operation and business. This should be aimed at assisting startups to sustain the initial years of business when they are unable to generate venture capital funds or have to sacrifice a large share of the business for early stage seed funds.

Proposed modules of the National AI Marketplace (NAIM)

It is proposed that the marketplace, called here as National AI Marketplace (NAIM), be divided into 3 distinct modules, at least in early stages of development. These are as follows:

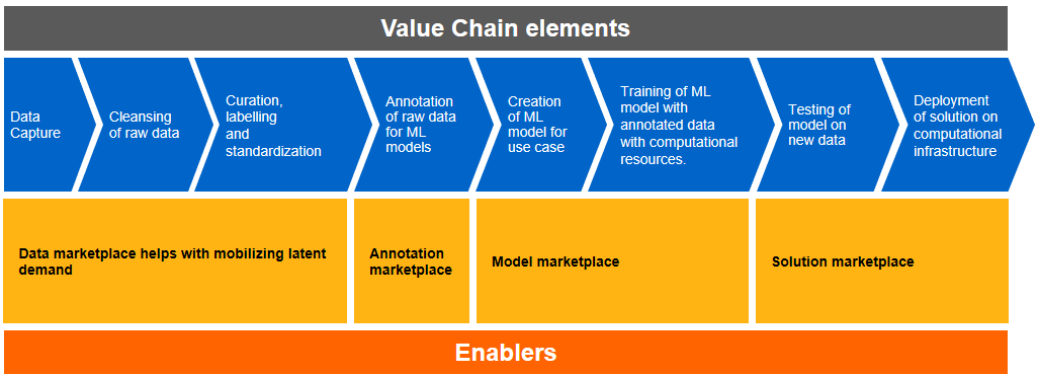
a) Data marketplace

b) Data annotation marketplace

c) Deployable model marketplace / Solutions marketplace

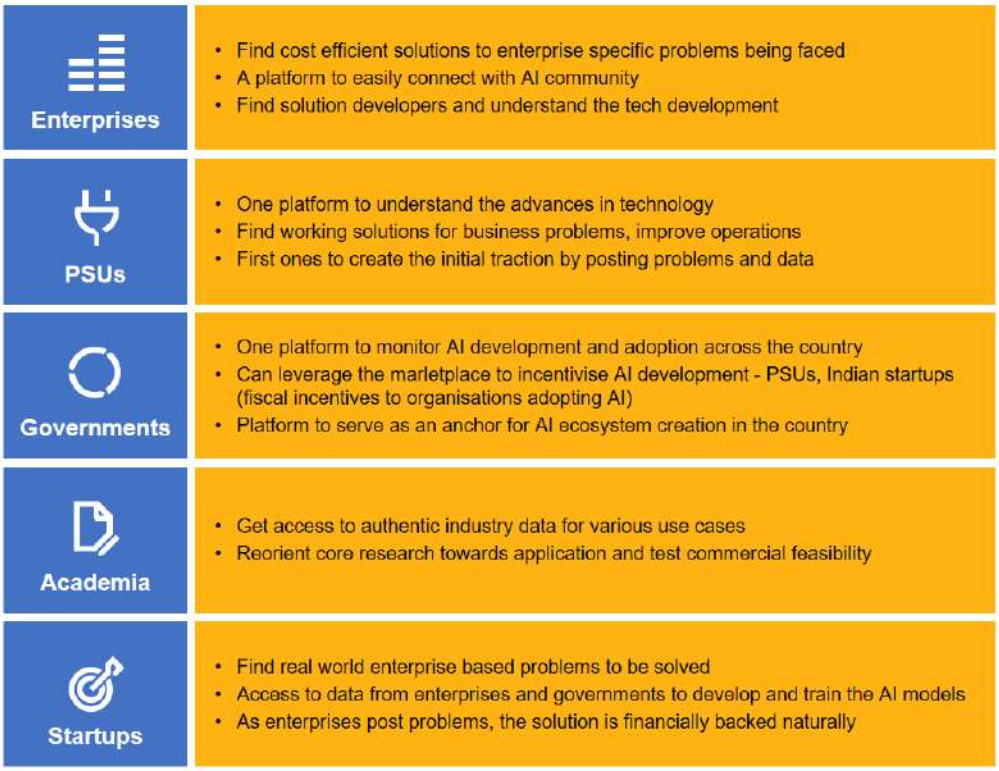
The AI development value chain, discussed above, is supported by these three modular marketplaces in terms of easing collaboration, reducing time and cost of collecting & annotating data, and bringing multiple solutions deployment at one place for scale and network effect.

Figure 29: *Enablers for AI value chain*



The marketplace mechanism being proposed here will be aimed at easing the adoption efforts of all participants – private enterprises, PSUs, governments, startups and academia. A common platform which brings together enterprises and AI solution builders will trigger off the initial collaboration towards building AI solutions and adopting them at scale.

Figure 30: Stakeholders for the proposed NAIM



Data marketplace

Today, incumbents continue to enjoy an oligopoly in building sustainable business models in AI for two main reasons: (a) they can successfully buy data in the informal market setting due to availability of resources and reach to negotiate one-time contracts continually, and (b) they have specialised departments to work on different facets of the development value chain. In the process of informal data acquisition as well, there is no proper price discovery mechanism in place. Compounded with the issue of a huge asymmetry of information in the sector between data providers and buyers, data providers are likely to be substantially underpaid.

This further compounds the problem of data access and thwarts the creation of a healthy marketplace as well as incentives. In a formal marketplace being proposed here, new entrants in the AI model creation / training space will find it easier to raise resources for buying data. This will be due to VCs having enough information to verify the funding demands and a lower risk resulting from a more level playing field and equitable access to new data sources.

While data has been compared to oil ad nauseam, there are a few differences which must be highlighted:

a) *Reusability:* Copied data is as valuable as the 'original'. Therefore, traceability of ownership is a challenge for a viable marketplace model where the originators of the data should get compensated from resale of the data.

b) *Compounding value:* Unlike oil, the value of a dataset increases not just with quantity but also structural complexity. For example, the value of a health dataset of a million individuals and an education dataset of the same individuals is more than the sum of its parts because of the value added by the connections between the different elements. Therefore, it is difficult to quantify the value of new datasets as it is dependent on the existing datasets one has and how one plans to connect the new data with the existing data.

c) *Permanence*: Unlike the oil, the inventory of data continually increases with time. However, this raises the complexity in transporting larger and larger datasets.

d) *Ubiquity*: Since data is ubiquitous and continually increasing in quantity, it follows that a way to automatically curate new datasets is needed so that any market does not overwhelm users with outdated or irrelevant data.

The fundamental properties of data and its collection outlined above create challenges which make the creation of a centralised data marketplace difficult. The primary assumption of a data marketplace in operation is that data custodians will share the data on the marketplace. Organisations may be reluctant to do so. Further, the operations of the data exchange, financially sustained by transaction fees and rigid pricing models, also add to inefficiencies and discourage data sharing.

Features of a viable data marketplace – how it can be implemented?

To incentivise a larger supply of AI training data sets and services, it is necessary to ensure availability of data, an audit trail mechanism to curb reselling of the same data, and ways to address security and privacy concerns. One solution is to have a centralised, trusted party to host the data on behalf of data providers and enforce the rules of the game. However, this is not practical for a variety of reasons, not the least because most data providers will not agree to let a replicable resource such as data be stored somewhere else, but also because of cost, control and trust concerns.

A more effective way to address these concerns is a decentralised data marketplace that is based on blockchain technology.

The exchange platform should have the following features for data providers to share data:

a) Traceability,
b) Access Controls,
c) Compliance with local and international regulations, and
d) Robust price discovery mechanism for data

The proposed data exchange marketplace will attract data providers and model builders / trainers to build AI products. The process of exchange, with enforced provisions of privacy and anonymisation, brings a market determined value to data and thus forces the existing informal data exchange economy, without any privacy protection, to move towards a formal economy.

The government can establish a committee of experts, researchers, AI developers and regulators to create the standards the data marketplace will adhere by and explore how can it be put in implementation.

The unique approach that one such initiative, i.e., Ocean Protocol has taken is to create what they call a Proofed Curation Market with two subcomponents: (a) Data exchange validation on blockchain, (b) Data curation market based on a derivative token. The former lays out a mechanism for data and value exchange, the latter enables the market to decide which datasets are important and creates an incentive for even free data to be put on the platform. Essentially, data consumers can bet on the viability of even free data and thereby earn money.

See *Appendix III* for more on data ecosystems as enablers.

Data annotation marketplace

Of the three marketplace modules defined, the data annotation marketplace is perhaps the most mature.

Given the large costs associated with the annotation exercise, a popular mechanism that has emerged is that of crowdsourcing using un-trained or non-expert anonymous annotators on the internet. Though this gives rise to its own set of challenges, it may be considered a solution to the problem of cheaply and quickly acquiring annotations for the purposes of constructing a variety of models. To understand the potential of crowdsourcing, consider the observation of a research paper in 2004[38]: a crowd of 5,000 people playing an appropriately designed computer game 24 hours a day, could be made to label all images on Google (425,000,000 images in 2005) in a matter of just 31 days.

However, some challenges that emerge in the annotation crowdsourcing process are as follows.

a) *Creation of verification mechanisms*: Dependence on untrained professionals gives rise to issues in ensuring high accuracy in annotation and thus potentially higher accuracy of the trained algorithm. Verification mechanisms such as multiple annotation, verification of samples by trained professionals, etc. are some of the ways to solve this problem.

b) *Establishing* verified traceability: Mechanisms must thus be put in place to identify poor annotators, and create mechanisms to reduce the effect of error generated.

Features of a data annotation marketplace – how it can be implemented?

Platforms such as Amazon's MTurks, Figure-Eight (previously known as CrowdFlower) or ClickWorker are perhaps the most popular crowd-sourced annotation platforms. Each allows for building of mechanisms of verification such as multiple annotations, creation of test cases, sample verification by professionals, etc. to increase accuracy of annotations.

Analysis indicates that most annotation tasks can be categorised on the basis of sensitivity to expertise of annotators. This is defined as the degree to which annotator expertise can affect the accuracy of annotation. The figure below considers how approaches to the annotation process could differ with changing nature of the task.

Figure 31: *Approaches to annotation*

Sensitivity to degree of expertise	Example of task	Motivation	Potential annotators	Verification mechanism
Low	Classification of object in presented image	GWAP, ML, ALTR	General population	Multiple annotations
High	Classification of medical scans	ML	College students, professionals,	Multiple annotations, Verification by professionals

GWAP = Games with a purpose, ML = Mechanised labor for financial incentives, ALTR = Altruistic motivation

Simple object classification or regional language translation are examples of tasks that have low barriers to annotator expertise. However, tasks such as classification of medical images, requires robust mechanisms to verify annotations, and also ensure *verified traceability* to identify poor annotators, and create mechanisms to reduce the effect of error generated.

In India, vast public datasets are available at organisations such as ISRO (India Space Research Organisation), ICAR (Indian Council of Agricultural Research), All India Radio (AIR) and NIC (National

[38] *Luis von Ahn and Laura Dabbish*, *2004*, *Labeling images with a computer game*

Informatics Center) among others, but are unfortunately largely unused due to lack of proper annotation, prohibitive cost associated with the process, and lack of proper mechanisms for acquisition. It is envisaged that these datasets be used to create initial corpora of datasets, and encourage enterprise action in the area. A crowdsourced platform for annotation, on the lines of the highly popular government's citizen participation platform *myGov.in*, can be established by the government (to be operated by a private player) where seed data for annotation can be provided by the organisations mentioned above.

Deployable Model Marketplace

Given the level of market for AI tools, AI-as-a-Service business model, and a plethora of standardised problems faced by enterprises at scale, a deployable model segment in the marketplace, which brings together the buyers and suppliers of AI solutions, can greatly ease and expedite the adoption of AI. It would give visibility to the existing solutions, address information asymmetry and generate awareness amongst the relevant constituents. This AI-as-a-Service model segment would further build upon and synergise with the previous two layers of the marketplace model viz. data and annotation.

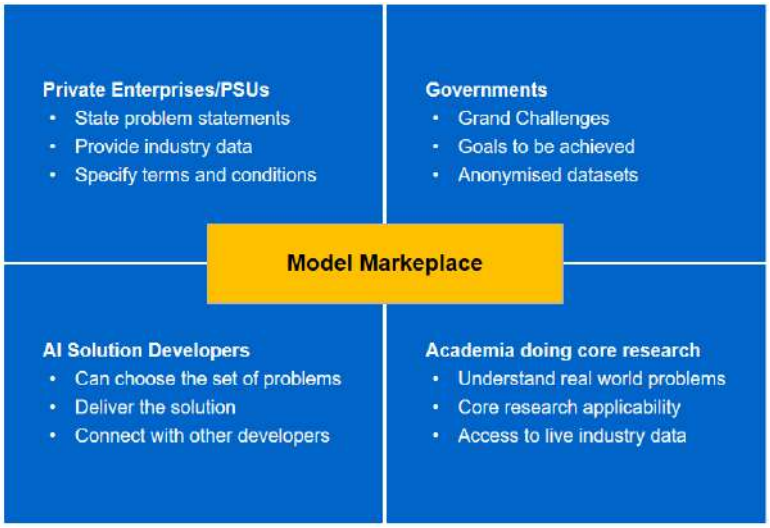
Many objectives aimed to be achieved by AI algorithms can now be categorised as standardised problems. The following are some of the standardised and ubiquitous problems most of organisations face in their activity streams:

a) object detection in images or video streams;
b) conversational smart chatbots (text and speech);
c) speech-to-text and text-to-speech;
d) assistive diagnostic solutions;
e) language recognition and transcription;
f) contextual data mining to discover complex patterns;
g) price optimisation;
h) data collection, curation and annotation for specific business use;
i) supply and demand forecasting; and
j) server, app and web uptime/downtime prediction.

Here models refer to the weights of the AI model trained on the data, which contains the predictive power. The inferencing of the model will require integration with other IT stacks and will be a customised process. Hence, this module of the marketplace should focus on defined deployable models that can be bought or sold. These models could either be bought and sold as data, or more likely, as a service or an API layer that can be charged per use.

The model marketplace will attract enterprises which can post problems which they are seeking to be solved using AI tools. Governments can be another party to announce specific problems which they intend to be solved. This demand side will be supported by the community of AI solution developers (consisting of startups, enterprises, individual researchers etc.) who will either showcase their existing developed and tested product suites, or customise their products as per the demand placed by enterprises. Academic researchers will be another major beneficiary of the marketplace as they will get valuable insights about problems faced by the industry and will get access to real industry data.

Figure 32: Stakeholders of NAIM



Challenges to adoption of a marketplace model for AI models are as follows:

a) *Ensuring quality:* Marketplaces require continuous monitoring and maintenance of quality. For example, Amazon enforces this on vendors, and this goes beyond user ratings. Continuous scrutiny of products, sellers and buyers will be required.

b) *Responsive product development required:* The correct set of features of this platform will have to be discovered through iteration and adaptation, since there isn't precedent of such AI platforms to draw from. The platform should allow for trial based access to models to allow for iterative product evaluation by buyers.

c) *Sophistication of PSUs for data sharing and evaluation against specs:* If data is to be provided over the platform, solution-seekers will need to know how to provide clean datasets, balance data to sufficiently represent infrequent occurrences, hide test data for evaluation of solutions while sharing training data, define appropriate specs, decide on appropriate evaluation methods, etc. The platform should be able to prescribe some basic standards for data sharing which will allow non-ML savvy PSUs to post problems and data.

d) *Incentives for seeding network effect:* The marketplace needs a certain critical mass of solution seekers and providers for gaining traction, to subsequently trigger network effects. This is true of all marketplaces. Incentives to attract both seekers and providers initially will need to be figured out. PSUs and government agencies may act as initial seeders/users of the platform with right incentives and mandates.

e) *Potentially rich feature set needed for adoption:* Such platforms can work well when there are systems like a credible rating system, payment escrow, search and sort to aid discovery, recommendation engines, etc. These will need to be established at NAIM also.

How can the stakeholders engage in NAIM?

NAIM, as a marketplace for enterprises, will encourage all business entities, government agencies, startups, AI researchers, system integrators and academic research institutions to sign up on the platform. Following may look like a snapshot of the marketplace where all the stakeholders engage in their respective activities on the platform:

Private / public sector enterprises

a) Hospitals: assistive diagnostics for physicians and radiologists.

b) Distribution companies: predict peak load, identify default prone customers.

c) Travel and tourism: predict peak traffic routes, pricing optimisation customers, offer alternative routes/modes for traffic optimization.

d) Large industrial or real estate establishments: smart power management, utilities management, real-time failure detection and analysis.

e) Cinema: automated subtitles in regional languages (speech-to-text and translation).

Government authorities

a) Traffic authorities and police: real-time object detection in CCTV video feeds, face detection and object tracking, number plate detection.

b) Citizen engagement: conversational chatbots (speech and text) in Indic languages for grievance redressal and query management.

c) Document analysis and management: text recognition and sentiment analysis from physical documents, analysis and summary.

d) All India Radio and Prasar Bharti: speech-to-text services for transcription and analysis.

e) Road and rail transport: face and eye movement detection (at the edge) for analysing driver fatigue and alarm signal, freight route optimization.

f) Agriculture: crop health and soil health analysis through remote sensing and IoT, farmer advisory services for input control and market demand prediction.

AI Solution Providers

a. India AI startups: solutions for natural language processing and Indic languages, diagnostic assistants, image recognition/object detection.

b. Industrial Internet-of-Things: capturing data from machines and providing insights in areas of energy management, retail and inventory management.

c. On Edge AI devices: CCTV video feed monitoring and warning signals.

d. Academic Research Institutions.

e. Centers of Excellence: access to industry data for various use cases, model training and evaluation.

Box 19: Why the focus on PSUs

Private enterprises run on the basis of market conditions and business drivers of revenues and profits. Due to either absence of markets or failure of efficient markets, many sectors of operation of PSUs are for achieving socio-economic goals of equality and distribution of wealth. There is not much incentive for PSUs for improving operational efficiency; they are further constrained by regulations imposed on publicly funded entities in activities of procurement, manpower hiring etc.

Technology adoption, especially AI, can significantly improve operational efficiency and drive down operational costs incurred by PSUs operating in financially unsustainable but socially critical areas. This

improvement in operational costs will make them financially more viable. The NAIM allows the public-sector firms to directly scout for solutions which they can use in their business operations by significantly easing their information search for solutions. Additionally, PSUs and government agencies can also be the first driver of the network of the NAIM where initial pool of problems and industry data is shared by them which will attract several solution developers, especially startups and researchers to sign up for the platform.

Mechanisms to realise marketplace model

It is proposed that a decentralised data marketplace based on distributed ledger technology be set up. In the future, such a platform can enable advanced privacy preserving AI techniques such as multi-party computation directly on individual pieces of data.

Instead of attempting to build such marketplaces, which it is not equipped to do so, the government should come out with enabling regulations, such that these marketplaces can be set up by private players. These would include regulations and standard permissions for any personal data being sold, standards for anonymisation, standards for ensuring annotation accuracy, and cybersecurity standards for the module dealing with deployable models.

An initial effort may be made by the government to build the NAIM platform where all relevant stakeholders can sign up. However, open market competition will allow other marketplace operators to come in with innovative services and up-to-date data on suppliers and buyers on the marketplace. NAIM will lead to a more engaging collaboration, quicker and easier data access and accelerated adoption of AI among enterprises and public authorities in the country.

Ethics, Privacy, Security and Artificial Intelligence

Towards a "Responsible AI"

AI is going to be the tipping point in technological evolution of mankind, with human dependence on machines and algorithms for decision making never been such deep. Thus, any strategy document on promoting AI, necessarily needs to be conscious of the probable factors of the AI ecosystem that may undermine ethical conduct, impinge on one's privacy and undermine the security protocol. Appropriate steps to mitigate these risks need to be an integral part of any such strategy.

While the issue of ethics would concern the biases that an AI system can propagate, the privacy concerns are largely on collection and inappropriate use of data for personal discrimination. Issue of security arises from the implications and the consequent accountability of any AI system.

While addressing the above issues, one needs to be conscious of the potential vulnerabilities of our extant regulatory and societal structures which are dependent on human judgment and control, and thus subject to inherent biases and discrimination. Thus, to say that extant decision making systems – individual, societal, regulatory or even judicial – are entirely devoid of these shortcomings would be a fallacy as these are dependent upon human limitations of knowledge, precedent, rationale and bias (explicit or subconscious). Delegation of some aspects of that decision making to algorithms, which may well be able to ingest and process many more parameters as compared to a human, may likely result in systems with reduced bias, discrimination and improved privacy protection. However, even if a technological intervention helps us delegate that responsibility to an algorithm with improved outcomes, it is extremely important that we set much higher standards for privacy and protection in case of AI tools.

Ethics and AI

Fairness / tackling the biases AI

Based on the premise that a large set of well-diversified data may be an accurate description of the world, most of the developer community takes a technocratic attitude that data-driven decision making is good and algorithms are neutral. However, this argument does not recognise the fact that the existing data may have biases, which may have got reinforced over time. The issue of fairness is at the forefront of discussion in academic, research and policy fora, and definitely merits a combined dialogue and sustained research to come to an acceptable resolution. One possible way to approach this would be to identify the in-built biases and assess their impact, and in turn find ways to reduce the bias. This reactive approach, use-case based, may help till the time we find techniques to bring neutrality to data feeding AI solutions, or build AI solutions that ensure neutrality despite inherent biases.

Transparency / opening the "Black Box"

Presently, most AI solutions suffer from what is commonly known as the "Black Box Phenomenon", with very little or no understanding of what happens in between and only the input data and results being the known factors. This is due to the reliance in most current AI systems to incrementally improve the performance as defined by a narrow set of parameters, with developer's emphasis being less on how the

algorithms are achieving the requisite success. However, calls for explaining the decision-making process will gain momentum as AI systems are increasingly relied upon for decision making that has significant consequences for a large section of population.

Opening the Black Box, assuming it is possible and useful at this stage (there is considerable debate on that as well), should not aim towards opening of code or technical disclosure – few clients of AI solutions would be sophisticated AI experts - but should rather aim at "*explainability*". With extended disclosure though, what needs to be balanced is whether the algorithm's parameter may induce the individuals and companies to change their behavior and in turn game the system. Clearly, more collaborative research is required in this area.

Box 19: Decoding Explainable AI

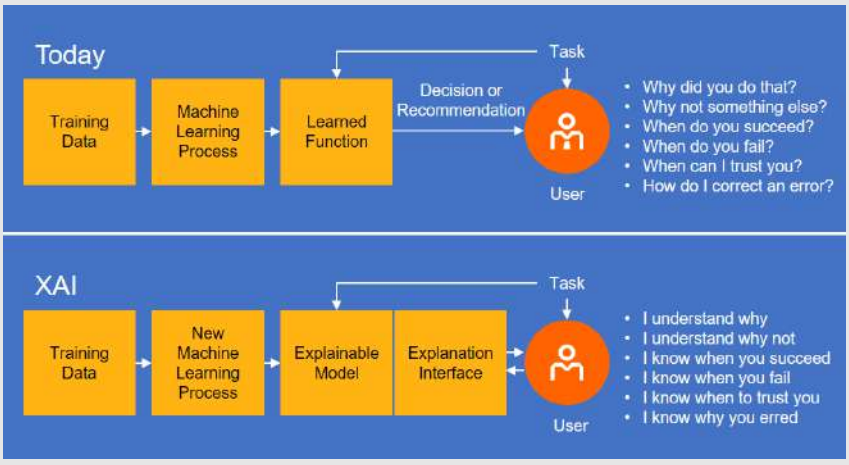
Explainable Artificial Intelligence (XAI) is an evolving area of research which has received a lot of attention from the research community and the broader society alike. XAI project by DARPA is one such project which has shown substantial progress in explaining how and why machine learning algorithms work in a certain way. The aim of the XAI project, as described by DARPA is:

The Explainable AI (XAI) program aims to create a suite of machine learning techniques that:

- Produce more explainable models, while maintaining a high level of learning performance; and
- Enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

The machine learning algorithms of tomorrow should have the built-in capability to explain their logic, enumerate their strengths and weaknesses and specify an understanding of their future behavior.

Figure 33: XAI Concept



XAI is one of a handful of current DARPA programs expected to enable "third-wave AI systems", where machines understand the context and environment in which they operate, and over time build underlying explanatory models that allow them to characterise real world phenomena.

In May 2018, XAI researchers are expected to demonstrate initial implementations of their explainable learning systems and conduct pilot studies of their Phase 1 evaluations. Full Phase 1 system evaluations are expected in November 2018.

***Credit**: DARPA Explainable Artificial Intelligence (XAI)*

Privacy and AI

AI models, solutions and their application depend on generation, collection and processing of large amounts of data on individual, entity and community behaviour. Data collection without proper consent, privacy of personal data, inherent selection biases and resultant risk of profiling and discrimination, and non-transparent nature of AI solutions are some of the issues requiring deliberation and proper recourse.

However, the current debate on data usage have two distinct aspects. Firstly, there are concerns that companies are harvesting significant amounts of consumer data and using it inappropriately to gain insights about consumers. Key here is that the consumer may not have access to these insights or the ability to derive value from them. Beyond compliance, companies can consider how to create awareness of how they use consumer information and the value they provide in return, which can build trust in their brand and services.

Secondly, there are concerns that companies are amassing large data sets and thereby building an unfair competitive advantage. Datasets themselves have little intrinsic value without the ability to extract meaning from them. The dataset is a necessary, but not the only, component of delivering meaningful insights from data. Having the tools to analyse it and the experience to understand its meaning are the others. While those that have access to large datasets and by the nature of their business models have data network effects, which enables them in turn to build a first mover advantage when it comes to perfecting their algorithms and driving business value, this does not necessarily negatively impact the consumer.

Dealing with privacy issues

a. *Establish a data protection framework with legal backing:* The work being done by Justice Srikrishna Committee on data protection law is very opportune and timely. The 7-core principles of data protection and privacy – informed consent, technology agnosticism, data controller accountability, data minimisation, holistic application, deterrent penalties and structured enforcement – are quite comprehensive and should provide a strong privacy protection regime in the country once enacted.

b. *Establish sectoral regulatory frameworks:* Apart from having a central privacy protection law, due to diverse and fast changing nature of the technology, sectoral regulatory frameworks may also act as additional protection to 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.

c. *Benchmark national data protection and privacy laws with international standards:* European Union's General Data Protection Regulation (GDPR) guidelines, which have been enforced in May 2018, encourage design of less-privacy invasive systems. French laws give a right to explanation for administrative algorithmic decisions, making it much more comprehensive than GDPR on administrative decisions. India's privacy protection regime will have to be continually updated to reflect understanding of new risks and their impact.

d. *Encourage AI developers to adhere to international standards:* Leaders and practitioners from across the world have come together to frame standards for safe and privacy preserving AI. The Global Initiative on Ethics of Autonomous and Intelligent Systems of the IEEE has a chapter on 'Personal Data and Individual Access Control in Ethically Aligned Design'. Indian enterprises and developers need to build these standards into AI design itself.

e. *Encourage self-regulation:* Data Privacy Impact Assessment Tools can be used by AI developers and enterprises adopting AI solutions to manage privacy risks.

f. *Invest and collaborate in privacy preserving AI research:* New mathematical models for preserving privacy are being researched upon where risks of data exploitation and personal identification (from an anonymised dataset) can be reduced by limiting information one can gain from released data, irrespective of amount of side information available otherwise. India should collaborate on areas of research like Differential Privacy, Privacy by Design, Safety-Critical AI and Multi-Party Computations which enable protection of privacy despite data sharing at a wide scale.

g. *Spread awareness:* Privacy has been termed as a fundamental right by the Supreme Court of India. The protection of this right with its multiple facets in a fast-changing technological environment will not just depend on State enforcement but by also making the citizens aware of their rights and how they can protect them. People often unknowingly give consent to sharing their data which they would not have ordinarily done had they known the purpose their data were being put to. There is an urgent need to spread awareness among the individuals about the importance of consent, ethics and privacy while dealing with technology. A pan-India campaign in multiple languages, and inclusion of privacy rights in school and college curriculum can serve as effective mass outreach mediums to spread awareness.

Box 20: Differential Privacy

A concept developed by Cynthia Dwork in 2006, Differential Privacy aims at preserving identifiable user information irrespective of any outside information the aggregator agency holds. The dilemma of giving personalised service to users based on their individualised preferences while at the same time ensuring that user is not uniquely identifiable, using the data collected or any other public data, is not solved by traditional privacy preserving methods of cryptography. Differential privacy describes a promise, made by a data holder, or curator, to a data subject: "You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available." It addresses the paradox of learning nothing about an individual while learning useful information about a population.

Security in AI

The accountability debate on AI, which in most of the cases today is aimed at ascertaining the liability, needs to be shifted to objectively identifying the component that failed and how to prevent that in the future. An analogy can be drawn to how the airlines have become a relatively safe industry today. Every accident has been elaborately investigated, and future course of action has been determined. Something similar is needed to ensure safe AI.

One possible framework that can be mooted involves the following components:

a. Negligence test for damages caused by AI software, as opposed to strict liability. This involves self-regulation by the stakeholders by conducting damage impact assessment at every stage of development of an AI model.

b. As an extension of the negligence test, safe harbours need to be formulated to insulate or limit liability so long as appropriate steps to design, test, monitor, and improve the AI product have been taken.

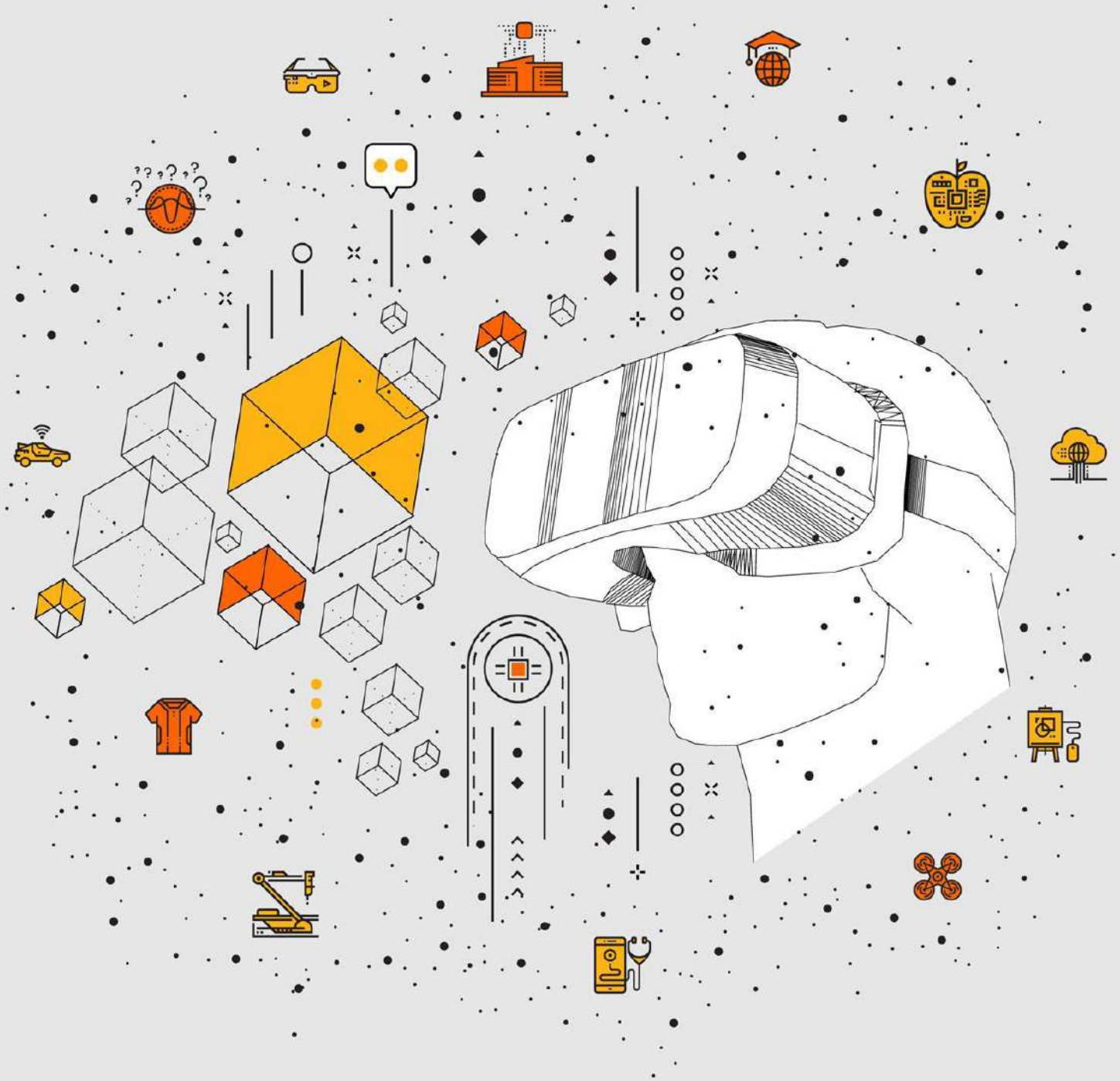
c. Framework for apportionment of damages need to be developed so that the involved parties bear proportionate liability, rather than joint and several liability, for harm caused by products in which

the AI is embedded, especially where the use of AI was unexpected, prohibited, or inconsistent with permitted use cases.

d. Actual harm requirements policy may be followed, so that a lawsuit cannot proceed based only on a speculative damage or a fear of future damages.

India can also take a leaf out of UK's playbook, where GBP9 million is being invested to establish a new Centre for Data Ethics and Innovation, aimed at enabling and ensuring ethical, safe and innovative uses of data, including AI. This will include engaging with industry to explore the possibilities of establishing data trusts to facilitate easy and secure sharing of data. A consortium of Ethics Councils at each Centre of Excellence may be set up to define the standard practice (on the lines of OpenAI charter). It would be expected that all Centres of Excellence adhere to standard practices while developing AI technology and products.

PART IV:
VISION AND ACTIONS FOR GOVERNMENT



Actions for the Government

Achieving the goal of #AIforAll requires long term and engaged institutional collaboration between all the stakeholders including the citizens. However, while playing the primary role in ensuring that this collaborative strategy succeeds, the government needs to be mindful of not crowding out the private sector. Role of the government thus needs to be one of a facilitator, an active promoter and wherever required, of an owner.

This section summarises the key recommendations, and the role of the government.

Figure 34: Government's role

Area	Recommendation	Government role
Research and Application		
Core Research	Setting up Centre of Research Excellence for AI (COREs)	Identify academic institutions, provide fiscal support to establish COREs focusing on core technology research in AI.
	PhD Scholarships	Institute National AI Fellowships to retain outgoing PhD students and attract researchers from foreign universities with attractive incentives and challenging projects.
	Inter-academia collaboration	Incentivise research collaboration between premier academic institutions through special grants while facilitating the formation of a global expert pool for core AI research.
	Faculty Fellowships	Provide Faculty Fellowships or Chairs in academic institutes to promote research in AI.
Applied Research	Setting up of International Centres for Transformational AI (ICTAIs)	Invite Expression of Interests (EoIs) from industry players to lead ICTAIs in various sectors (health, education, agriculture, smart mobility and smart cities), in collaboration with the government and academia. Build governance structure, provide fiscal support, formulate an IP model for ICTAIs and set up the ICTAIs under a PPP model through "challenge method".
	Setting up ICTAI Inc., overarching entity for ICTAIs	Establish "ICTAI Inc." as either society / section 8 company, with initial contribution from government and private sector representation, to select and fund ICTAIs.
Common Compute Platform	Setting up AI Research, Analytics and knoWledge Assimilation platform (AIRAWAT)	Set up a common cloud platform for Big Data Analytics and Assimilation with a large, power-optimised AI Computing infrastructure connecting all COREs, ICTAIs and other academic institutions with National Knowledge Network.
Intellectual Property	Building an attractive IP regime for AI innovation	Set up a task force, comprising jointly of Ministry of Corporate Affairs and DIPP, to examine and

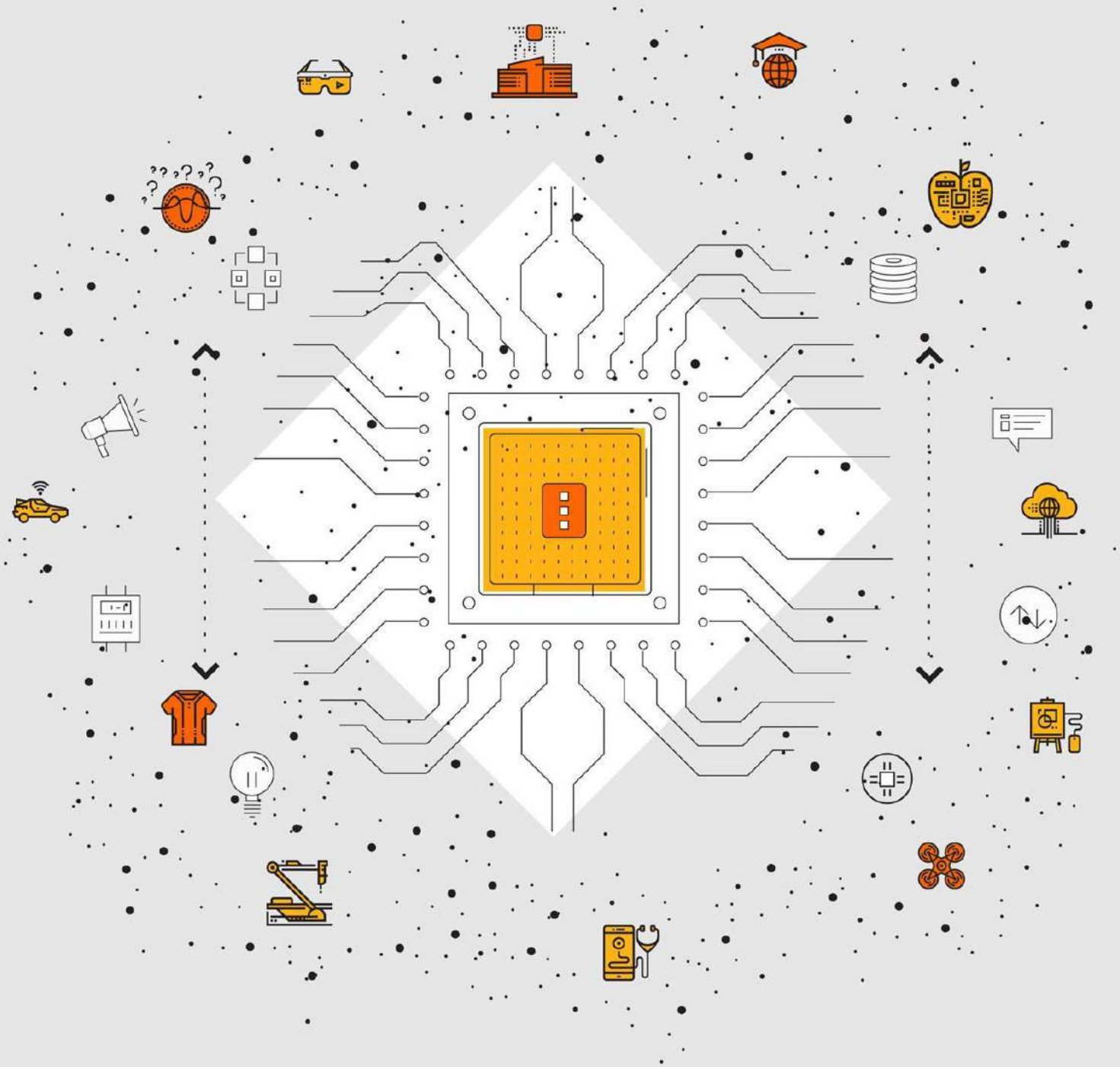
		issue appropriate modifications to the IP regulatory regime pertaining to AI.
Supra-national collaboration	Setting up CERN for AI	Take the lead in bringing together the relevant parties to create People's AI, the CERN for AI – national governments, industry, academia and international community of researchers.
Reskilling and Training		
Workforce	Promote formation of future service sector jobs	Incentivise creation of service sector jobs of the future such as data annotation through tax holidays or inclusion in CSR activities.
	Recognition and standardisation of informal training	Set up AI / Data Science training standards, as per National Skills Qualification Framework, and provide certifications to training institutes.
	Promote employee reskilling	Incentivise investment in training of employees through tax breaks and grants for employers.
Colleges	Expansion of quality education in data science and AI	Incentivise colleges / universities to adopt credit-bearing MOOCs in their curriculum.
	Promote cross-disciplinary AI education	Introduce Bridge Courses in AI for post-graduates in non-computer science or data science domains.
Schools	Introducing AI / ML in schools	Introduce AI modules in Atal Tinkering Labs.
Overall	Continuously assess the changing nature of jobs	Constitute a standing committee or taskforce to examine and report on changes in employment induced by adoption AI.
Accelerating Adoption of AI		
Data Sharing	Opening up government datasets	Establish platforms for making datasets in the area of social sector (either collected during implementation of a scheme or in normal business processes) available for open public use in a machine readable form.
Data Annotation	Creating and making India specific annotated datasets public (on the lines of ImageNet)	1. Catalyse partnerships with the various academic institutions and public / private agencies in making annotated India specific data available for advancing AI research. 2. Explore partnerships and co-fund building of large corpora of data across domains, as a means of laying the foundation for startups and enterprises to build applications and services tailor-made to the Indian context.
Crowdsourcing Annotation	Annotation of data – images, text, speech etc. via crowdsourcing	Announce grand challenge tasks for tagging of images, text or videos, and devising reward based mechanisms through data market place to aggregate the content from the various participating members.
Nation-wide adoption	Enabling a multi-stakeholder owned and	Create governance guidelines, explore partnerships and co-fund the establishment of: *1. Data marketplace*

	managed National AI Marketplace	2. *Data annotation marketplace* 3. *Deployable model marketplace* to develop the data supply ecosystem, ease collaboration, reduce time and cost of collecting & annotating data, and bring multiple solutions deployment at one place for scale and network effect.
Visibility in Collaboration	Making information search for collaborations easier	Set up an AI Database portal for easy dissemination of information on projects being implemented via collaboration among government-academia-industry-researchers-startups to enable resource matching.
Awareness and Adoption in Government	Making decision makers aware about transformative potential of AI	1. Workshops, live demonstrations, 2. AI Readiness Index to highlight best practices across states, and 3. Create Central-State shared fund for AI led development projects to be taken up by States.
Government and PSUs as seeders for network effect	Making governments and PSUs leaders in adoption of social AI tools	Help create a pipeline of AI research projects for the COREs, ICTAIs through grand challenges to be given by the government and PSUs. Incentivise public agencies to adopt and employ AI in delivering service through financial support; extra budgets for R&D; tax incentives and awards.
Partnerships and Collaboration	Industry – Academia – Trade Bodies – Venture Capital Collaboration	Encourage close collaboration between industry, academia, trade bodies and venture capital to implement "AI+X" paradigm.
Startup Support	Support systems for AI based startups	Establish incubation hubs and venture funds specifically for AI startups in collaboration with State Governments.
Responsible AI Development		
Ethical and Responsible Research in AI	Making COREs and ICTAIs adopt ethical practices	Set up a consortium of Ethics Councils at each CORE and ICTAI to define the standard practices and monitor their adoption.
Privacy and Security	Instituting a data privacy legal framework	Address and implement data protection framework, which protects human rights and privacy without stifling innovation in India.
	Creating sectoral regulatory guidelines	Collaborate with industry to come out with sector specific guidelines on privacy, security and ethics – on manufacturing, financial services, identity, telecommunication, robotics etc.
	Collaborating on privacy preserving technology research in AI	Support COREs to do research in new mathematical models and technology for preserving privacy; encourage international collaboration.
Sustainable Research	Setting up Centre for Studies on Technological Sustainability (CSTS)	Set up CSTS to address issues relating to ethics, privacy, legal aspects, social sustainability and global competitiveness of the technologies developed.

Financial implication of recommendations

A greater understanding of the financial implications of the recommendations made in this report will be realised after consultations with the different stakeholders. Given the emphasis on research, skilling and creation of the ecosystem, the recommendations would involve a significant budgetary allocation by the government. This would be in addition to funds being provided as part of the Digital India and Startup India initiatives.

APPENDIX

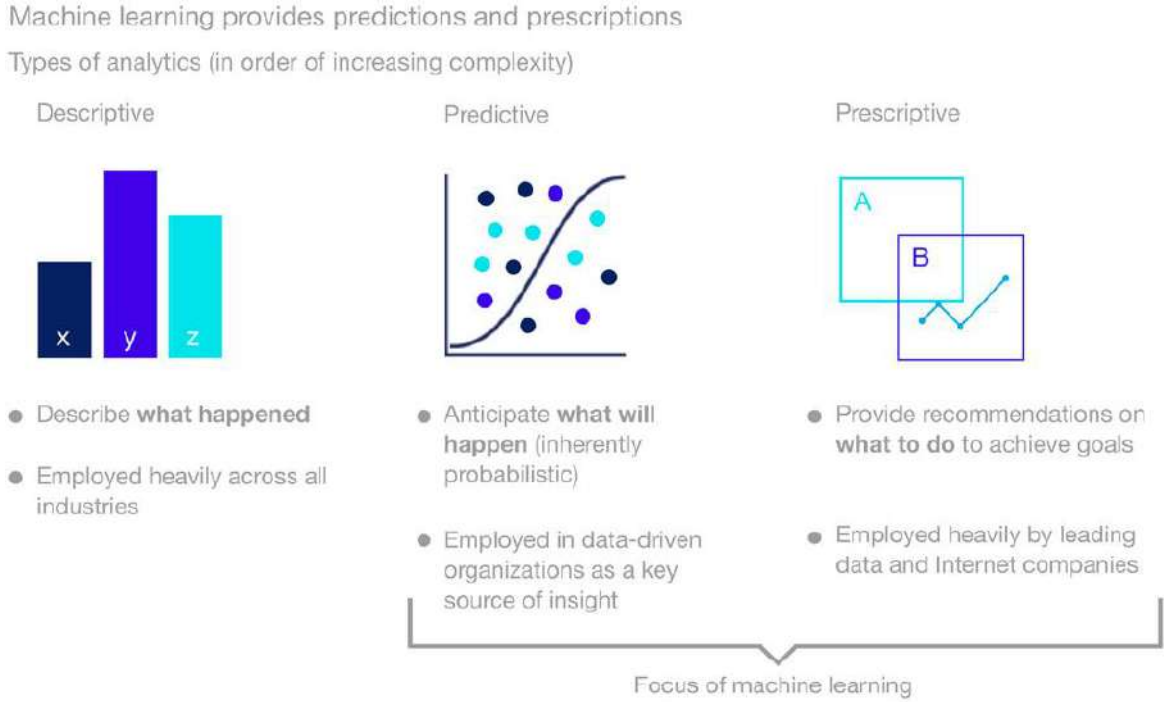


Appendix I: Artificial Intelligence Explained

An Executive Guide to Artificial Intelligence[39]

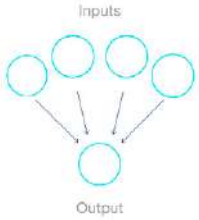
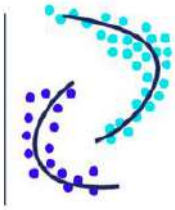
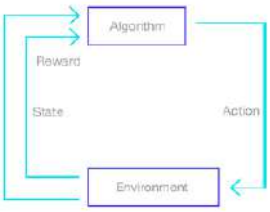
Machine Learning: a definition

Most recent advances in AI have been achieved by applying machine learning to very large data sets. Machine learning algorithms detect patterns and learn how to make predictions and recommendations by processing data and experiences, rather than by receiving explicit programming instruction. The algorithms also adapt in response to new data and experiences to improve efficacy over time.



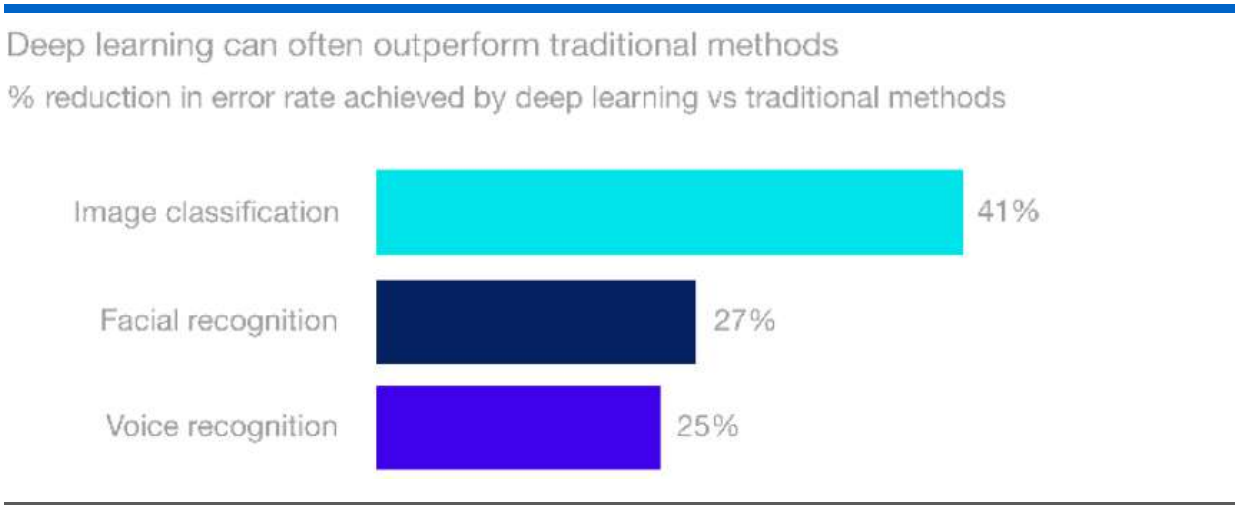
[39] *Replicated from McKinsey Analytics An Executive Guide to AI*

Understanding the major types of machine learning

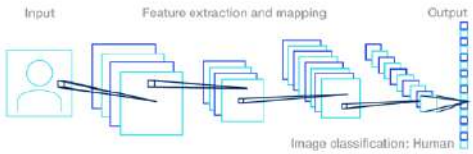
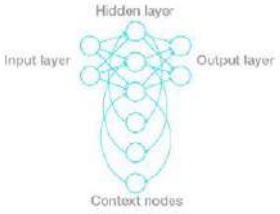
	Supervised Learning	Unsupervised Learning	Reinforcement Learning
			
What it is?	An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output (e.g., how the inputs "time of year" and "interest rates" predict housing prices)	An algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns)	An algorithm learns to perform a task simply by trying to maximise rewards it receives for its actions (e.g., maximises points it receives for increasing returns of an investment portfolio)
When to use it?	You know how to classify the input data and the type of behavior you want to predict, but you need the algorithm to calculate it for you on new data	You do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you	You don't have a lot of training data; you cannot clearly define the ideal end state; or the only way to learn about the environment is to interact with it
How it works?	1. A human labels every element of the input data (e.g., in the case of predicting housing prices, labels the input data as "time of year," "interest rates," etc.) and defines the output variable (e.g., housing prices) 2. The algorithm is trained on the data to find the connection between the input variables and the output 3. Once training is complete – typically when the algorithm is sufficiently accurate – the algorithm is applied to new data	1. The algorithm receives unlabeled data (e.g., a set of data describing customer journeys on a website) 2. It infers a structure from the data 3. The algorithm identifies groups of data that exhibit similar behavior (e.g., forms clusters of customers that exhibit similar buying behaviors)	1. The algorithm takes an action on the environment (e.g., makes a trade in a financial portfolio) 2. It receives a reward if the action brings the machine a step closer to maximising the total rewards available (e.g., the highest total return on the portfolio) 3. The algorithm optimises for the best series of actions by correcting itself over time

Deep Learning: a definition

Deep learning is a type of machine learning that can process a wider range of data resources, requires less data preprocessing by humans, and can often produce more accurate results than traditional machine learning approaches. In deep learning, interconnected layers of software-based calculators known as “neurons” form a neural network. The network can ingest vast amounts of input data and process them through multiple layers that learn increasingly complex features of the data at each layer. The network can then make a determination about the data, learn if its determination is correct, and use what it has learned to make determinations about new data. For example, once it learns what an object looks like, it can recognise the object in a new image.



Understanding the major types of deep learning

	Convolutional neural network	Recurrent neural network
		
What it is?	A multilayered neural network with a special architecture designed to extract increasingly complex features of the data at each layer to determine the output	A multilayered neural network that can store information in context nodes, allowing it to learn data sequences and output a number or another sequence
When to use it?	When you have an unstructured data set (e.g., images) and you need to infer information from it	When you are working with time-series data or sequences (e.g., audio recordings or text)
How it works?	**Processing an image** 1. The convolutional neural network (CNN) receives an image – for example, of the letter "A" – that it processes as a collection of pixels 2. In the hidden, inner layers of the model, it identifies unique features, for example, the individual lines that make up "A" 3. The CNN can now classify a different image as the letter "A" if it finds in it the unique features previously identified as making up the letter	**Predicting the next word in the sentence "Are you free …..?"** 1. A recurrent neural network (RNN) neuron receives a command that indicates the start of a sentence 2. The neuron receives the word "Are" and then outputs a vector of numbers that feeds back into the neuron to help it "remember" that it received "Are" (and that it received it first). The same process occurs when it receives "you" and "free," with the state of the neuron updating upon receiving each word 3. After receiving "free," the neuron assigns a probability to every word in the English vocabulary that could complete the sentence. If trained well, the RNN will assign the word "tomorrow" one of the highest probabilities and will choose it to complete the sentence

Appendix II: Global Country Strategy Review[40]

What is happening around the world in AI?

Countries around the world are becoming increasingly aware of the potential economic and social benefits of developing and applying AI. For example, China and U.K. estimate that 26% and 10% of their GDPs respectively in 2030 will be sourced from AI-related activities and businesses. There have been tremendous activity concerning AI policy positions and the development of an AI ecosystem in different countries over the last 18 to 24 month – the USA published its AI report in December 2016; France published the AI strategy in January 2017 followed by a detailed policy document in March 2018; Japan released a document in March 2017; China published the AI strategy in July 2017; and the U.K. released its industrial strategy in November 2017.

Governments are reviewing and developing their position on the following areas to rapidly grow AI ecosystems:

a) Trigger demand in socially relevant sectors / segments

b) Gear up the supply side to fulfil demand – infrastructure (including data ecosystem, data stacks, high speed computing, etc.), talent, research

c) Set up an enabling system—governance, funding, partnerships

Trigger demand in socially relevant sectors

Different countries have identified different focus areas for AI development and deployment:

a) the USA: areas of interest include economic prosperity, educational opportunities, quality of life, national and homeland security. The USA is focusing on growing the AI ecosystem through public spending on contracts, e.g., the US Department of Defense spent over USD2.4 billion on AI-related technology in 2017 (2x increase from 2015).

b) China: areas of interest include education, healthcare, energy, transport, quality of life, city planning /IoT / robotics. China is focusing on developing and using AI for delivery of public services through financial support, developing talent pipeline, and leveraging international cooperation.

c) Japan: areas of interest include industrial productivity improvement, healthcare, medical care and welfare, mobility and information security. Japan is focusing on moving from the "Industry 4.0" paradigm to "Society 5.0" through the development of AI use cases for delivering public services.

d) France: areas of interest include healthcare, environment, transport mobility, defence-security. The government is planning to support AI startups through data availability, public spending and talent reskilling.

[40] *McKinsey*

e) U.K.: areas of interest include services, life sciences, agriculture and public-sector applications. The government is focused on growing innovative tech firms and making deals with the private sector to solve AI use-cases for delivery of public services.

Gear up the supply side to fulfil demand

Infrastructure

Most governments have taken the following two action steps, in varying degrees of engagement, to upgrade infrastructure and build a data ecosystem:

a) Creation of a data-solutions marketplace

b) Invest in upgrading computing infrastructure, 5G networks, etc.

Many countries have been reviewing a range of initiatives to facilitate creation of these marketplaces and upgrading computing infrastructure as well as connectivity. For example:

a) U.K. is exploring the feasibility of creating data trusts where the process of data sharing and storage is underwritten by the government. There is also a focus on defining data rights for potential participants on the platform. There are plans to invest GBP1 billion to upgrade digital infrastructure including rolling out 5G and full-fibre networks.

b) Japan has announced expanding its R&D tax exemption to include AI and big data as well as subsidies for building new robots with integrated AI. Its focus is on developing sector-specific platforms for public and private development of AI, followed by interlinking of different platforms to create an integrated AI ecosystem.

c) France is trying to streamline its innovation track with "innovation sandboxes" which would provide an open platform for innovation and offer resources for use in field-testing, etc. AI research institutes would have supercomputers specifically designed for AI usage and devoted to researchers and their economic partners during their shared initiatives.

d) China is focused on developing open source innovation platforms in partnership with private players like Baidu, Alibaba, and Tencent. Funding is available for 5G networks to enable "*intelligentisation*" and deployment of supercomputers, high performance semiconductor chips for AI use.

e) the USA is facilitating the creation of open source software libraries and toolkits, e.g., Open NLP, Weka toolkit, etc.

Talent

Countries are also significantly increasing allocation of resources for STEM talent development through investment in universities, mandating new courses (e.g. AI and law), and offering schemes to retrain people.

a) U.K. is planning to increase its R&D spend to 2.7% of its GDP by 2027, investing GBP42 million in teacher development and GBP64 million in the retraining scheme including digital training. They are planning to make it easier for Tier-1 applicants working on AI subjects to obtain work permits and have a path to residency. They are also planning to build over 1,000 government supported PhD institutions by 2025 and set up a Turing fellowship to support an initial cohort of AI fellows.

b) France is trying to triple the number of AI graduates in three years by offering new courses and doubling the starting salary of researchers in public institutions. They also want to attract talent from across the world by offering substantial salary hikes, supporting improvements in the quality of life and reducing administrative formalities. Further, France and the US have labs and panels to assess the impact of AI on the workforce.

c) U.S. is planning a USD200 million grant for STEM education focusing on computer science matched by a USD300 million industry grant.

d) Japan has convened a "*national consultative body*" with 3 universities and the Japan business foundation to develop education programs for reskilling.

e) China has launched a five-year university program to train at least 500 teachers and 5,000 students working on AI technologies. The program is a collaboration between government bodies, private companies and universities including Sinovation Ventures. China has set in motion a plan to develop 50 world-class teaching institutes and research institutions, 50 national-level high quality online open courses and 50 AI faculties by 2020 as part of the "AI+X" program.

Research

Universities and research institutions from the USA, China and Japan have led the publication volume on AI research topics between 2010 and 2016. In the US, Carnegie Mellon University (CMU), Massachusetts Institute of Technology (MIT) and Stanford are the top three universities in mean count of papers published across AI, systems, theory and interdisciplinary areas[41]. These universities have been pioneers of AI research in the US and have more than 100 faculty members across different areas of AI research.

a) CMU has one of the oldest AI programs in the world – it was also one of the first to offer an undergraduate program. It has started the CMU AI program which is a collaboration forum for faculties across seven departments to work on multidisciplinary AI topics.

b) MIT has launched the Intelligence Quest to discover the foundations of human intelligence and its application to develop technology and tools.

c) Stanford has a AI4ALL program to increase diversity in AI research and education.

These three universities also feature on the top of any list on infrastructure and industry relations, e.g., IBM's Watson was developed in research collaboration with CMU:

[41] *CSRankings.org*

Research and course-work	▪ **100+ faculty members across departments** ▪ **1000+ students** engaged in AI research and education ▪ Launched **undergrad degree in AI** ▪ **IBM's Watson**, World champion robot soccer players, etc.	▪ CS & AI lab has **1200+ people** working on **900+ projects** ▪ Research ranges from **algorithm and theory** to applied **AI and machine learning** ▪ Launched **MIT Intelligence quest** to discover the foundations of human intelligence and develop tools	▪ **80+ faculty** across **17 research focus** areas ▪ **Andrew Ng** is an adjunct associate professor for Machine learning ▪ **AI4ALL programme** to increase diversity in the field of AI research
Research facilities	▪ State of the art **robotics institute** ▪ **AI stack** ▪ **Design for AI lab**	▪ Computational genomics lab ▪ **Distributed robotics lab** ▪ **Vision lab**	▪ Stanford AI lab – Toyota center for AI research
Research partnerships	▪ U.S. department of defence ▪ NASA ▪ Amazon ▪ Google ▪ Microsoft ▪ Oculus	▪ $240 million investment from IBM for 10 years ▪ 60+ member companies across telcos, finance, consumer electronics	▪ JD.com ▪ Google ▪ DiDi ▪ Panasonic ▪ UST Global ▪ Tencent ▪ Samsung

Source: Press search

Chinese universities have also established extensive research partnerships with Baidu, Alibaba and Tencent:

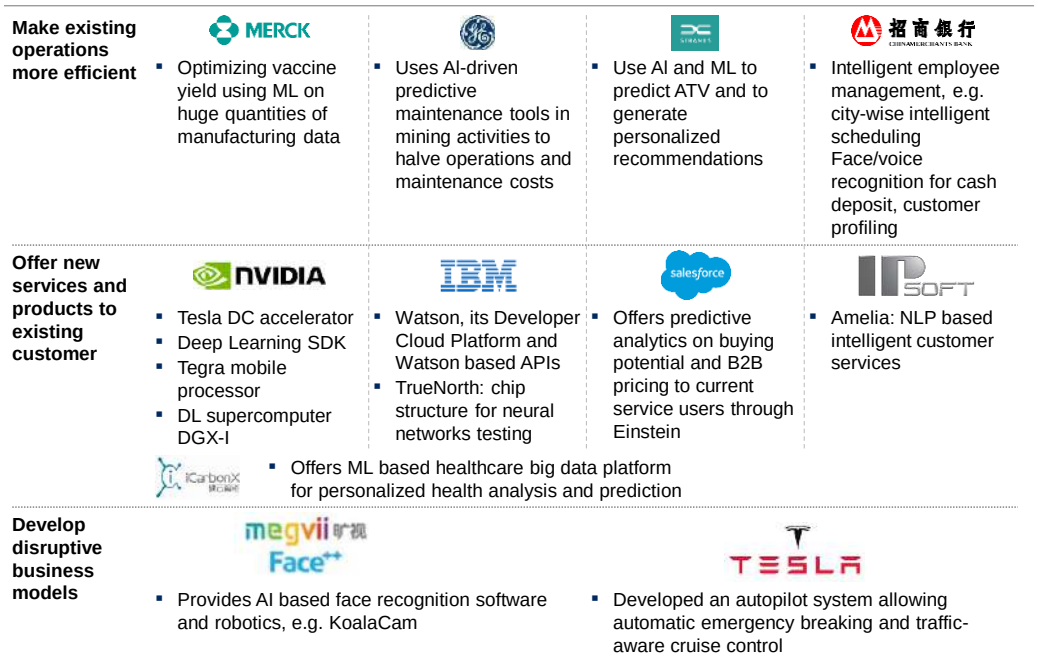
a) Baidu has announced a USD104 million partnership with the Peking University to further research in AI-related topics including information science and medicine.

b) An AI centre for law and legal issues was also unveiled at the Peking University to research on applications to improve legal efficiency while decreasing burden on judges.

c) Alibaba's Aliyun is working with the National Engineering laboratory on big data systems and software at the Tsinghua University.

Japanese research has historically been hardware centric with robotics as one of the major focus areas of development. With the growing demand for AI, academia in Japan is rapidly reorienting itself to theoretical and applied research in the sector. Riken Centre for Advanced Intelligence Project and Advanced Industrial Science and Technology are the nodal research institutes for industry collaboration.

With AI led disruption of multiple industries, it is imperative for traditional industry players to increase speed and agility of insight generation, design and digitise customer journeys as well as develop efficiency in delivering journey transformations. Consequently, traditional players like GE and Merck are investing heavily to power their offering through AI interventions.



Source: Press search

Set up an enabling system

Governance

Most of the governments have established / utilised existing centralised umbrella body for budgetary planning of AI interventions and for formulating strategy and drafting policies. The National Science and Technology Council in the USA, Strategic Council for AI technologies in Japan, AI council in the U.K. are nodal agencies for planning and designing AI initiatives. These central bodies typically consist of ministers, representatives from industry and nominated members from academia, e.g., the UAE has a Minister of State for AI. Similarly, for implementation and delivery of AI initiatives:

a) U.K. has a dedicated department "Office of AI" to collaborate with multiple departments, ministries and other stakeholders to deliver AI projects

b) France has a shared specialist centre of 30 members to help provide specific inputs and implement projects in other departments

c) In China and Japan, individual ministries and departments are responsible for implementing AI solutions across different sectors. For example,

   i. China: National Development and Research Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, Central Military-Civil Fusion Development Commission Office, the Central Military Commission (CMC) Science and Technology Commission, and the CMC Equipment Development Department, etc.

   ii. Japan: Ministry of Health, Labour and Welfare, Ministry of Land, Infrastructure, Transport and Tourism, and the Ministry of Agriculture, Forestry and Fisheries, etc.

While due importance is given to the central planning and implementation entities, the role of local governments in solving area-specific challenges through application of AI is becoming increasingly important. For example,

a) London has a Smart City board and a Chief Digital Officer to apply best practices in smart infrastructure and AI.

b) Over 19 Chinese cities, including Beijing, Shanghai, Hangzhou, Zhejiang, Tianjin, have been mandated to develop their own city-level AI agenda.

Utilising AI-related solutions at the grass-roots level to solve real local challenges has the potential to truly democratise its use.

Funding

Governments are significantly increasing the funding for the AI ecosystem.

a) Apart from the increase in R&D spend to 2.7% of its GDP, U.K. has created a GBP725 million industrial strategy challenge fund, GBP1.7 billion for transforming cities fund, reforming enterprise investment scheme and venture capital trusts to unlock GBP7 billion over 10 years. Additionally, they have instituted a GBP2.5 billion investment fund at the British Business Bank to incubate tech startups and reforming rules for pension funds to mandate inclusion of AI in their investment portfolio.

b) Japan is planning to increase its science and innovation budget by JPY900 billion by 2020 for AI. Different Japanese ministries are also funding R&D centres, e.g., Ministry of Economy, Trade and Industry (METI) is funding R&D centres at the National Institute of Advanced Industrial Science and Technology (AIST).

c) France is planning to spend EUR1.3 billion to develop AI-led interventions.

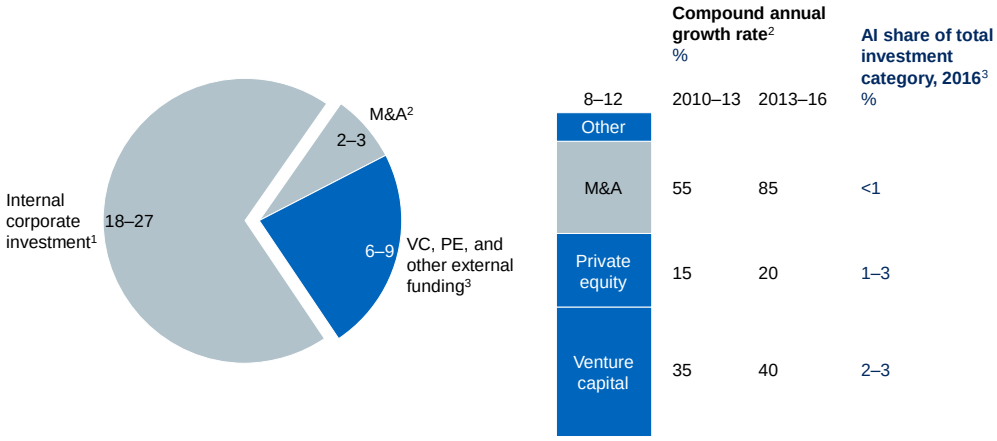
d) China is funding a massive growth in network infrastructure and creating megaprojects.

e) the USA is increasing its spend on AI-related contracts with the Department of Defense alone spending USD2.4 billion. Other large spenders include departments of agriculture, veteran affairs and homeland security.

Private companies have dominated the investment in AI with internal corporate investments as the biggest mode of spend. Other large sources of funding for startups within the sector include venture capital (VC) and private equity (PE). The two biggest areas funded are machine learning and computer vision, followed by natural language processing, autonomous vehicles and smart robotics.

Technology giants dominate investment in AI

Investment in AI, 2016[1]
$ billion



1 Estimate of 2016 spend by corporations to develop and deploy AI-based products. Calculated for top 35 high tech and advanced manufacturing companies investing in AI. Estimate is based on the ratio of AI spend to total revenue calculated for a subset of the 35 companies.
2 VC value is an estimate of VC investment in companies primarily focused on AI. PE value is an estimate of PE investment in AI-related companies. M&A value is an estimate of AI deals done by corporations. "Other" refers to grants and seed fund investments. Includes only disclosed data available in databases, and assumes that all registered deals were completed within the year of transaction. Compound annual growth rate values rounded.
3 M&A and PE deals expressed by volume; VC deals expressed by value.

Source: McKinsey Global Institute report on 'Artificial Intelligence and the next digital frontier'

Digital native companies in the USA and China are exploring new areas for the development and application of AI to solve customer-centric use-cases. These range from developing smart AI game systems like Google's AlphaGo, to virtual assistants like Amazon's Alexa and Apple's Siri. There is also a push to develop open source AI platforms like Google's TensorFlow to Baidu Brain. Alibaba has also partnered with the Malaysian government to launch the first smart city AI platform outside China (Hangzhou was the first example). Apart from internal incubation and development, these tech giants and VCs are also investing heavily in startups focused on AI.

	Google	amazon	Baidu 百度	Alibaba.com
Make existing operations more efficient	▪ Customized **search results** ▪ Customized **ad ranking** ▪ **Google cloud** platform	▪ Developed **Amazon web services** for cloud computing ▪ **Customer services** ▪ Warehouse management	▪ Customized search results and ad ranking ▪ 020: **order prioritization and route planning**	▪ Leverages AI to improve **operational efficiency** and **customer experience** of e-commerce ecosystem
Offer new services and products to existing customer	▪ Provides **open source AI platform** Tensorflow ▪ Offers **customized ASIC** for ML TPU ▪ Online translation ▪ **Graphic** and **voice search**	▪ Integrates AI into cloud services offering ▪ Provides mobile robotic based **automated storage and retrieval systems**	▪ Opens AI technology platform **Baidu Brain** ▪ Offers AI solutions in **finance, healthcare**, and **traffic & transportation** industries ▪ Netflix like streaming service **iQiyi**	▪ Launches AI data apps construction platform ▪ Provides **enterprise level AI solutions** in with Ali-cloud as foundation ▪ **Tmall genie** – voice controlled smart home assistant
Develop disruptive business models	▪ Virtual assistant: **Google Now** ▪ Game AI system: **AlphaGO** ▪ **Self-driving car** system ▪ Smart home: **Google Home**	▪ **Virtual assistant** – Alexa ▪ Polly – **life-like speech** ▪ Rekognition – **image analysis** ▪ Lex – **conversational engine**	▪ DuerOS: **Virtual assistant** ▪ Little fish – **voice controlled family robot** ▪ **Apollo** – set of artificial intelligence driven tools for **self-driving vehicles**	▪ **Smart city AI platform** developed for Kuala Lumpur ▪ **Smartmesh connectivity solution** supports many-to-many bluetooth mesh technology

Source: Press search

Partnerships

These countries are also leveraging different combinations of public-private-academia to develop and promote AI:

a) In the U.K., a public-private-academia partnership was established as "sector deals" to improve productivity. The expansion of tech parks through the tech nation program is also an example of public-private partnership. The government is also trying to develop regional R&D partnerships between universities, large corporations and investors in the sector, e.g., BT has partnered with 15 universities across the U.K. on creating AI powered, next generation data infrastructure.

b) Japan has instituted new programs to triple research-industry collaboration by 2025 (including co-locating industry employees with researchers). They have also signed collaboration pacts with the US and Israel for technology transfer and joint R&D projects.

c) Japan is trying to foster solutioning of challenges faced by large corporation by connecting them with startups, e.g., Japan Open Innovation Council, New Energy and Industrial Technology Development Organisation (NEDO) pitch, etc. to connect startups with corporations.

d) China has formed a "national team" with large private players including Baidu and Tencent to undertake fundamental and applied research across different AI topics, e.g., Baidu is working with the Chinese government to develop brain-inspired intelligent technology

Meanwhile, recent developments in the digital ecosystem have triggered a discussion on implications for regulations on data protection and privacy. EU has released a comprehensive legal framework for data protection called General Data Protection Regulation (GDPR). This framework details the rights of individuals (consent, data portability, etc.), obligations of businesses (define and share how they will use personal data, norms for data processing, data protection impact assessment, etc.) and plan of action in case of a data breach (data breach notifications, compensation to individuals, penalties, etc.). As AI grows

rapidly across geographies and sectors, governments across the world are actively working on developing data privacy and security regulations.

Furthermore, governments are playing an active role in developing AI ecosystems to capitalise on the social, economic benefits and establish leadership in the field of AI.

a) *Social benefits*: Governments are focusing on sectors ranging from education to healthcare, agriculture to transport mobility with a view to significantly improve quality of life of its citizens.

b) *Economic benefits*: Governments have defined substantial economic aspirations through development and implementation of AI. While China aims to grow AI's contribution to GDP to 26 percent and the U.K. by 10 percent by 2030, Japan has estimated the economic impact of AI application at JPY 1.1 trillion by 2045.

c) *Leadership in AI*: Given the rapid pace at which AI technology is evolving, governments are setting themselves up for success with support from the private sector and academia. However, the models of engagement vary depending on starting points, challenges and appetite for public funding and regulation.

If some countries decide to wait for a few years to establish an AI strategy and put in place the foundations for developing the AI ecosystem, it seems unlikely that they would be able to attain and match up to the current momentum in the rapidly changing socio-economic environment. Therefore, the need of the hour is to develop a policy framework that will help set up a vibrant AI ecosystem in India.

The following table highlights funding commitments made by governments across the world to promote AI research and application:

Country	Area	Funding
Belgium	• AI research in academia	• Two funding agencies – FWO (Flanders) and FNRS (Wallonia). • FNRS spent approximately EUR1.8 million per year in the period 2011-2017 and FWO approximately EUR6.7 million per year. • Between 2011 and 2017 around 67 out of 241 AI-related applications (representing 2.3% of all applications) submitted to FNRS were funded, and 175 out of the 832 AI-related applications sent to FWO were also accepted.
China	• AI startups	• In China, governments play a deliberate and explicit role in funding scientific research (giving USD800,000 to USD1 million in subsidies to AI companies).
Denmark	• AI startups	• The Innovation Fund Denmark has provided EUR20 million as funding for big data in 2017.
Germany	• AI basic research • Applied AI research	• With an annual budget of more than EUR3 billion, the German Research Foundation (DFG) is the main source of funding for basic research in AI in Germany. • In the past thirty years (1988-2018), applied AI has been funded continuously by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung - BMBF), for a total of EUR215 million. • Current annual investment in AI is EUR40-50 million. Between 1988 and 2017, DFKI received EUR200 million from BMBF. There is additional funding allocated to universities and other research centres by the government.

		• BMBF and the Federal Ministry of the Economy and Energy (Bundesministerium für Wirtschaft und Energie - BMWi) are currently funding a selection of Industry 4.0 projects such as Mixed Reality Production 4.0, for a total of EUR550 million since 2013. • The program Smart Services World II aims to address areas where digitalisation could have an impact for the economy. It has funding of EUR50 million from BMWi (2017-2021).
Ireland	• AI startups	• The Irish government spends, through the Irish Economic Development Agency (IDA), Enterprise Ireland, and Science Foundation Ireland, over EUR700 million on R&D annually. Enterprise Ireland funds Irish companies and is the largest VC fund in Europe.
Israel	• AI ecosystem – partnerships	• The Israeli government has several grant funding schemes for promoting collaboration and knowledge transfer between academia and industry, such as Magnet and Magneton. • The Israeli Science Foundation has a rich history of funding AI projects in academia that provide researchers with a high degree of freedom in their research compared to other countries.
Netherlands	• AI research – academia	• The main funding body for academia in the Netherlands is the Netherlands Organisation for Scientific Research (NWO). Since 2002, NWO has funded 119 research projects containing the term 'artificial intelligence' and 142 research projects containing the term 'machine learning'. In 2015, a programme on 'Natural Artificial Intelligence' was launched, which has funded five projects.
Spain	• AI in industry	• Since 2016, EUR170 million has been invested in the Industry 4.0 project Industria Conectada 4.0 under the National R&D and Innovation Plan. Industry 4.0 focuses on skills, cooperation, industrial adoption, and digital technologies (robotics, AI, cloud, cybersecurity, big data).
Sweden	• AI research – academic • AI ecosystem - partnerships	• The Research Institutes of Sweden is currently setting up an AI centre (RISE AI), with an initial turnover of SEK50 million (approx. EUR4.9 million) per year in R&D, over 4 startups, more than 50 experts and around 30 active industrial collaborations (e.g. Nokia, Ericsson, ABB, and H&M). • Vinnova, Sweden's Innovation Agency, has funded 190 AI-projects totaling SEK398 million (approx. EUR38.9 million) in the past 6 years.

Appendix III: Data Ecosystem

A key enabler

The true value of AI will not be found in, say, an algorithm or a neural network itself.

Today, the leading algorithms are available as software packages either commercially or open source. Storage infrastructure and huge computing resources has become commoditised and easily available from various vendors in the market. The challenge still remains with data – access to high quality, reliable data along with appropriate mark-ups still remains a challenge.

Understanding key issues related to data thus becomes crucial while evolving the national AI strategy.

Access to vast quantities of data is vital for AI to be effective. Large platforms and technology companies who create monoliths of high volumes of data have a distinct advantage compared to smaller companies and startups, thus leading to a very skewed market.

Convergence of data is another challenge, with much data being either “dark” (unstructured, not readily usable) or disparate (hard to combine). Organisations need to be able to converge and make sense of data from sources such as Internet of Things (IoT) sensors and social networks. The greater the data density and variety, the greater are the chances of finding “unknown unknowns” – relationships that were not known to exist or were not looked for at all.

Data annotation is crucial, given the requirement of tagged or annotated data required for machine learning or AI. In most cases, data needs to be annotated manually or in semi-automated ways for the purpose of machine learning, even though sometimes annotated data can be generated automatically from the source. Technological advancements in data annotation systems are not at a level where manual annotation can be replaced.

Box 21: ImageNet

The ImageNet database has led to huge advancement in the field of image recognition. The ImageNet database was presented for the first time as a poster at the 2009 Conference on Computer Vision and Pattern Recognition (CVPR) in Florida by researchers from the Computer Science department at Princeton University. The ImageNet project is a large visual database designed for use in visual object recognition software research. Over 14 million URLs of images have been hand annotated by ImageNet to indicate what objects are pictured; in at least 1 million of the images, bounding boxes are also provided. ImageNet contains over 20,000 ambiguous categories; a typical category, such as "balloon" or "strawberry", contains several hundred images. The database of annotations of third-party image URLs is freely available directly from ImageNet; however, the actual images are not owned by ImageNet.

The 2010s saw dramatic progress in image processing. In 2011, a good ILSVRC classification error rate was 25%. In 2012, a deep convolutional neural net achieved 16%; in the next couple of years, error rates fell to a few percent. With the 2012 breakthrough "combined pieces that were all there before", the dramatic quantitative improvement marked the start of an industry wide AI boom. By 2015, researchers reported that software exceeded human ability at the narrow ILSVRC tasks. However, as one of the challenge's organisers, Olga Russakovsky, pointed out in 2015, the programs only have to identify

images as belonging to one of a thousand categories; humans can recognise a larger number of categories, and also (unlike the programs) can judge the context of an image.

Overcoming issues related to data access

a) *Government data sharing*: Government of India has large amounts of data lying in silos across ministries. The government can launch a mission of making all these data available for public good after undertaking proper privacy checks. For example – climate data, non-strategic remote sensing data, regional language speech (from All India Radio), soil health data etc.

b) *Corporate data sharing*: Corporates based in India may be mandated to share their data for social good. For example, sharing transportation pattern of individuals / mass transits, collected by service providers and aggregators, can help the city planners help in planning routes, predicting and managing traffic.

c) Consent based data sharing: Lot of the data about individuals are personal in nature and hence cannot be shared by third party who have the access to these data – like financial institutions or hospitals. Upon proper and informed consent from the citizens, these anonymised data may be shared for the purpose of artificial intelligence and data analytics.

d) *Digitised and crowdsourced collection of data by government*: Huge amounts of money and time is spend every few years to carry out the household consumption survey. A mechanism, as adopted by online social networks, to incentivise individuals to share details of their consumption pattern via an app can greatly reduce the cost of manual surveys and lend itself to big data analysis and AI applicability.

Often AI based transformation follows as a next step to digital transformation – in that case, the data for the purpose of learning exists with the organisation. For example, a financial institution would have access to large amounts of historic customer records to develop the AI system for predictive loan defaults.

People or organisations outside the data ecosystem might have limited access to this data. For example, a new company entering in the space of rural financing would need access to huge amounts of customer data along with information on the default rates to develop the risk based lending model. To bridge this access gap, some of these datasets are made available by the research communities as shown in the tables below.

Datasets for Computer Vision and Image Processing

Dataset Name	Dataset Size	Publisher	Publisher Type	Country
MNIST	60,000	United States Census Bureau	Institute and Bureau	USA
CIFAR 10 & CIFAR 100	60,000	Canadian Institute for Advance Research	Research Institute	Canada
ImageNet	14,000,000	Princeton University	University	USA
LSUN	10,000	Princeton University	University	USA
Pascal Voc	5,00,000	Oxford University	University	UK
SVHN	6,00,000	Stanford University	University	USA
MS COCO	2,50,000	Google, CMU, etc.	Research Consortium	USA
Visual Genome	108,000	Stanford University	University	

Labeled Faces In The Wild	13,000	University of Massachusetts	University	USA

Datasets for Natural Language Processing and Text Mining

Dataset Name	Dataset Size	Organisation	Org. Type	Country
TREC	120K-3.6M	NIST Information Technology Laboratory	Research	USA
WikiText	100 million	Salesforce	Company	USA
Question Pairs	4,00,000	Kaggle	Company	USA
SQuAD:	1,00,000	Stanford	University	USA
CMU Q/A Dataset:		Carnegie Mellon University	University	USA
Maluuba Datasets:	120K	Microsoft	Company	USA
Billion Words:	1 Billion	Google	Company	USA
Common Crawl:	1.81 billion	Common Crawl	Non-Profit Organisation	USA
bAbi:	Multiple	Facebook AI Research (FAIR)	Company	USA
Stanford Sentiment Treebank:	107,785	Stanford	University	USA
20 Newsgroups:	20,000	CMU & UCI	University	USA
Reuters	21578	AT&T Labs	Research	USA
IMDB	25,000	Amazon	Company	USA
UCI's Spambase	4601	UCI	University	USA

It is clearly evident that universities and research laboratories in the USA in particular, have been the leaders in sharing these datasets, which has contributed significantly in developing a strong AI research base in the USA. The Government of India can also play a crucial role, working with the various academic institutions, in making annotated India specific data available for advancing AI research.

Appendix IV: What Do the Markets Say?

Approaches to evaluating focus sector areas

One way of evaluating which sectors to focus on is to follow the money i.e. which sectors have seen the most amount of VC funding in AI space. Venture Capital funding, in a sense, is a long-term view on a technology / solution, given that the median time to exit for most VC investments is more than 8 years.

The AI ecosystem is essentially based on 5 pillars:

a) policy makers,

b) large companies,

c) startups,

d) universities and

e) multi-stakeholder partnerships.

Taking into account this perspective as well, sectors that are most pursued by VCs could be a good proxy for sectors to be focused on, for well-funded startups are capable of pushing the technology frontiers and bringing ambitious solutions to fruition.

AI remains the most active industry vertical in VC funding landscape, with an aggregate total investment of more than USD31 billion globally across more than 3,600 disclosed deals in past 5 years[42]. Healthcare has been the hottest area of AI startup investments, with USD2.5 billion in VC investments in last 5 years, much of which has been fueled by medical imaging and diagnostics companies. Other active healthcare sub-segments include clinical trials & drug discovery and insights and risk analytics. The healthcare AI space has also been dominated by strong public partnerships in diagnostics: nVIDIA and GE, Google Deep Mind and NHS, AliHealth and AstraZeneca.

Another way to evaluate this could be to see where the large global tech companies (Google, Apple, Facebook, Amazon and Microsoft, popularly termed *GAFAM*, as well as Alibaba and its peers Baidu, Tencent and Xiaomi, popularly termed *BATX* in China) are most invested in. Among GAFAM, looking at the mentions of “Machine Learning” in respective earning transcripts since 2013, Google emerges as the company that has been highlighting its progress in AI / ML more than others.

Tracking Google’s AI / ML initiatives, it appears that the Mountain View based company is heavily invested in Healthcare. Indeed, GV, formerly Google Ventures, the venture capital investment arm of Alphabet Inc., has tripled its investments in Healthcare deals in 2017 compared to 2013. Google’s approach to using AI to tackle diseases and lifestyle management is based on three pillar approach :

a) Data generation: digitisation and consolidation of data through wearables and medical imaging etc.

[42] *CB Insights: “Artificial Intelligence Trends To Watch In 2018” and “Up And Up: Healthcare AI Startups See Record Deals”*

b) Disease detection

c) Disease / lifestyle management

Google is currently heavily focused on eye diseases (diabetic retinopathy), diabetes (detection and management), heart disease (including hearth condition monitoring), Parkinson's disease and multiple sclerosis. Google's AI algorithm for diabetic retinopathy, trained on 128,000 images, was on-par with a panel of ophthalmologists. Other areas that Google may be exploring include Chronic Lower Respiratory Disease, several types of cancer, mental and behavioral health and aging.

Google is also invested in powering the healthcare data infrastructure, as evident in its USD625 million acquisition of Apigee, which is building healthcare APIs catering to the latest health records interoperability protocols. Similarly, Deep Mind is building a data infrastructure to enabling building of apps that can analyse different data elements. Furthermore, Google is also building health data streams that third parties could integrate in their research. Another one of Google's interesting foray in healthcare AI is developing tools for doctors that are designed to augment their expertise.