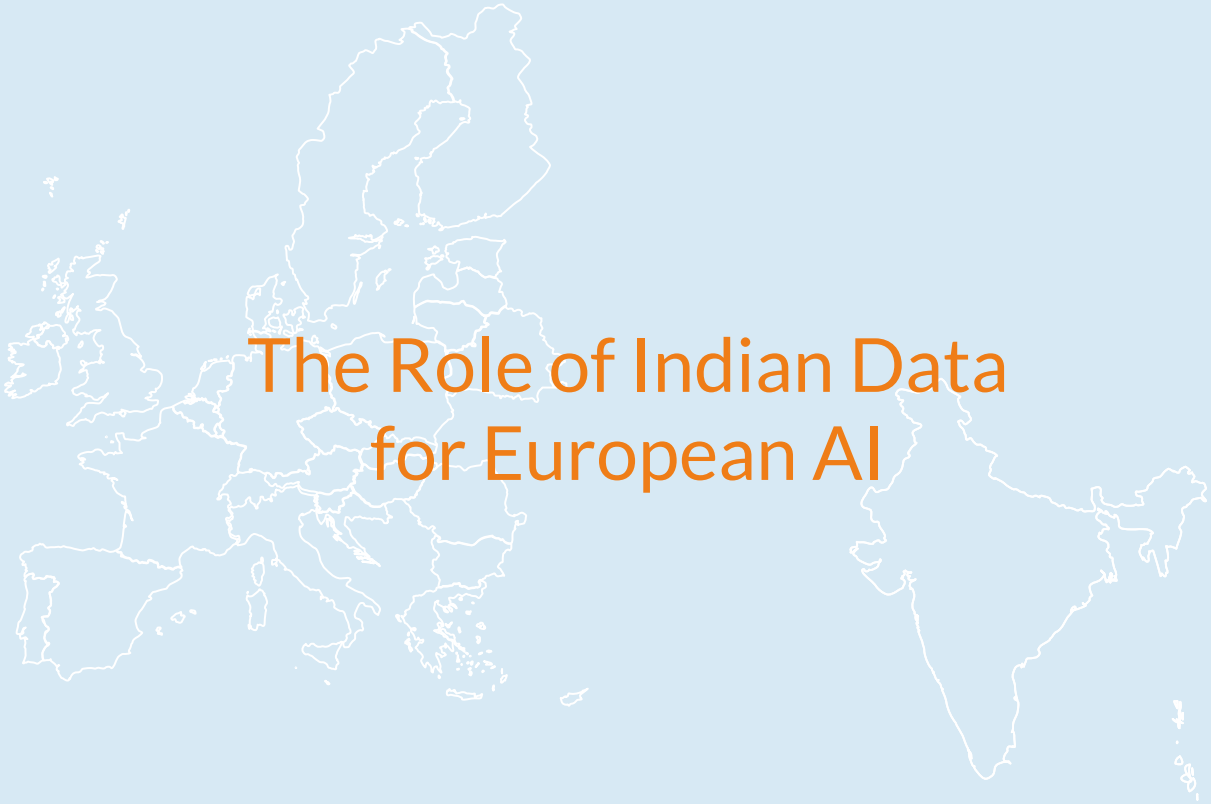
The Role of Indian Data for European AI

BertelsmannStiftung



The Role of Indian Data for European AI

Abstract

The study “The Role of Indian Data for European AI” investigates the potential benefits of a closer collaboration on data exchange between India and Europe. It explores the Indian AI landscape, Germany’s need for access to larger data pools, and requirements to make a cooperation possible. While no quick wins are to be expected from a closer cooperation, the long-term prospects are promising, provided that several important obstacles are cleared first. India’s new regulation on data privacy and security, due to be passed soon, appears to be a make-or-break issue.

contents

Contents

1. Executive Summary

After a period of sluggish progress in the Euro-Indian strategic partnership, the EU-India summit held in July 2020 reaffirmed that both sides are keen to deepen their cooperation in security, the environment, innovation and public health. The two large democratic blocs, together accounting for roughly a quarter of the world's population and GDP, see huge potential in closer cooperation. One area that seems particularly promising is innovation and artificial intelligence (AI). The leaders of India and Germany also put AI cooperation prominently on their agenda included in the joint statement after their last intergovernmental consultations in November 2019.

This push also builds on the assumption that data from India might be valuable to promote AI development in Europe. As India is a vast and diverse country, it produces a lot of data that might complement data available in Europe, and vice versa. But is this assumption valid?

Acknowledging the strategic importance of close cooperation between the EU and India, this study, prepared for the Bertelsmann Stiftung by CPC Analytics, an Indo-German AI consultancy, investigates whether a cooperation on data between India and Germany/the EU can boost the German and European AI ecosystem. It also asks whether a realistic case can be made for AI cooperation, and especially for cross-border data exchange between India and Germany/the EU.

The study focuses on the situation in India to understand the potential for a data collaboration based on the actor and regulatory landscape in the country. The AI landscape in India is sketched in a first step and this knowledge is then complemented by interviews with relevant actors in the field in order to shed light on the practitioner's perspective of using cross-border data to build AI. As a regulatory environment allowing for data exchange is a necessary condition for such an exchange to happen on a large scale, the current regulatory situation and plans are described.

Indeed, the analyses of India's digitalization efforts and of the private sector landscape point towards a rapidly evolving environment for AI. India's government has launched a series of initiatives to support the country's digital development. Useful data is increasingly being collected and the country certainly has the talent pool and industry to make use of this data. These factors make cooperation on data for AI between Germany and India promising. Yet major obstacles are present, including regulatory and data availability issues. These obstacles are discussed in the study, which also takes a detailed look at two sectors, health and e-commerce, finding that collaboration possibilities vary greatly between them.

The potential for cooperation is considerable but remains theoretical for now with regards to large-scale data exchanges (i.e. across a wide variety of sectors and companies). A quick realization, let alone quick economic benefits, does not seem realistic at the moment. This situation is unlikely to change before there is clarity on the effects of India's Personal Data Protection (PDP) bill, which is not expected until 2022.

The US and China can be regarded as the world's most advanced countries in the field of AI, both technologically and in terms of data access, while the rest of the world is following at a distance. Europe needs to find its place in this environment and make sure it is not left behind. Thus, on February 19, 2020, the European Commission published its European Strategy for Data, which lists as an important goal increasing access to larger data sets relevant to AI development. The long-term potential for cooperation with India in this regard is considerable for various reasons. Besides obvious factors like the country's large population, market size and a thriving IT services sector, three aspects of its digital economy stand out: the establishment of a digital baseline infrastructure ("India Stack"), the political push towards digitization (national AI strategy, work on PDP) and the digital start-up landscape, which has created more than 20 unicorns in recent years. All of these aspects result in a growing pool of data that could be used for AI.

While our study estimates the current size of the Indo-German AI-related economy to be roughly between €500 million and €1.5 billion (with total bilateral trade at about €20 billion), many obstacles need to be overcome before meaningful cooperation can be realized. Moreover, there seems to be "no low-hanging fruit" in Indian data that could be picked easily. At the moment, most data in India seems to be either privately held and inaccessible, context specific and not transferable, or of lower quality and thus not usable. India is making progress in this regard, however, and cooperation is likely to be promising for both countries in the longer run as more data of higher quality becomes available. A lot will depend on the realization and implementation of data protection laws in India and on whether a data-adequacy decision by the European Commission can be attained. Adding to these difficulties, the Hindu-nationalist agenda of the current Indian government has called the country's traditional role as a "natural partner" for Germany/the EU into question. Various ethical considerations will thus need to be resolved, too, before real progress can be made. Having argued so far from a European perspective, it should also not be forgotten that, while there is interest in India to cooperate more with Germany and the EU, the latter are not the first priority for India's very entrepreneurial and dynamic AI companies.

Partly because of the huge potential and partly because both countries will need strong partners in the future since they cannot become relevant counterweights to China or the US on their own, closer cooperation seems appropriate. The current geopolitical situation represents a window of opportunity in this regard. There is obvious potential for cooperation between India and Germany/the EU in the area of AI, but rather surprisingly the study could not identify major examples of cross-border data exchange for AI development. Even within companies, such examples seemed to be very limited.

Against this background, making a major political effort to foster cooperation in data for AI might not be the most promising candidate for inclusion on the bilateral agenda for now, as it would bind a lot of policy makers' time and resources for a long period and for a mission that is theoretically appealing but not yet fully proven in practical terms. Championing an Indian PDP bill that is likely to gain adequacy status in the EU thus really seems like the single most important high-level topic that policy makers should focus on for now. There is an open window of opportunity as the Indian data protection regulation has not yet been finalized and European initiatives might fall on fertile ground. If this succeeds, it will be worthwhile to deepen efforts for more cooperation further.

2. Introduction: The need for cross-border data partnerships for AI

2.1. Relevance of data access as the foundation for AI innovation

Artificial intelligence (AI) is expected to be a general-purpose technology that will affect almost all aspects of economic activity and transform social life.[1] Next to computational power, a key ingredient of current AI applications is access to vast amounts of data.[a]

Having access to a dataset that fits the demanding requirements of the current AI-workhorse technology, machine learning, allows firms to reap the benefits of this new technology and obtain a competitive advantage over rival firms. Rapidly accessing such datasets appears to be crucial, because the initial dataset used to train models is more important than subsequent data for achieving a breakthrough in AI.[2] At the same time, accumulating more relational data appears to be important, because it allows the innovating firm to improve productivity and market share (e.g. through economies of scale / scope of data).[3]

The approaches to achieving this differ: Large Silicon Valley-based firms have succeeded by quickly rolling out standardized services capturing (and defending) large market shares across the world – based on mobility of data. Chinese firms have built their databases mostly from their vast domestic (and highly competitive) market activity, where liberal sharing of data between the public and private sectors have allowed firms to innovate quickly, as have innovation-friendly government policies.[4]

Data accumulation in combination with the industrialization of learning have turned the digital transformation of economies into an "arms race" between firms, in which the winner takes the most.[5] In subsequent extrapolations of these observations from businesses to the country level, the debate predicts a similar competition between countries, in which some "countries are leading the data economy"[6] and "AI Superpowers" are shaping "a new data world order."[7]

Europe has tended towards a different framing. While the EU strategy for AI and the member states' national strategies have also made the competitiveness of their economies a key theme,[8] there has been a consistent emphasis on building AI "that respects the Union's values and fundamental rights as well as ethical principles such as accountability and transparency."[9]

Many European companies have been less rapid in digitalizing their business models and in adopting AI as an instrument to exploit the data. A "European/German model" of AI-driven innovation is still in the making.[10] Easing barriers to cross-border data transfer among EU member states is one element of the EU Commission's strategy for making "Europe fit for the digital future." For example, providing access to larger data sets relevant to AI development is a crucial problem that must be overcome, as outlined in the European Strategy for Data presented by the European Commission in February 2020 ("single European data space").[11, 12]

a For the purpose of this study, "data" is defined broadly as digitized information originating from people (e.g. through surveys or observing online behavior) or from machines (e.g. in production or supply chain processes).

2.2. Data access and sharing beyond national borders and the EU

From an economic theory point of view, increased data sharing between countries (and firms) can contribute to long-term economic growth. This observation is rooted in the non-rivalry characteristic of data, i.e. data can potentially be used by many people simultaneously without its utility being diminished for anybody.[b] If wide access to data is granted, more data is available to AI developers. In combination with other production factors (e.g. labor), data can then lead to increasing returns to scale, i.e. data accumulation can deliver long-term economic growth.[13]

From a practical perspective, data partnerships with non-European/non-Western countries represent a crucial complementary step to EU-wide data sharing. Engaging with partners from various contexts and getting access to this diverse data represents a technical imperative, because most AI systems are created from "learning-by-doing" and eventually need context-specific data to be robust.[14] Expanding data access to other non-EU countries would also determine the competitiveness of German/European companies in the future: New product offerings in the digital economy are often spun from data collected as a side product from other business activities.[15]

Developing a common understanding on cross-border data transfers for AI also offers a strategic opportunity: As technology policies become elevated to a geopolitical level (e.g. data protection and trade agreements, 5G, internet governance), international collaboration on cross-border transfers represents one aspect of influencing the future direction of digital governance on a global level.

b Of course, it needs to be acknowledged that non-rivalry – in combination with the ability to at least partially exclude others from using collected data – might lead firms to fear creative destruction from other people using the data and challenging the firm's current market position. Not only will these dynamics negatively impact the sharing of data, they can ultimately lead to harmful effects on competition (e.g. dominant market positions of large e-commerce firms). See sources at the end of the paragraph.

2.3. Germany's need to increase the data pool for AI development

"The internet is uncharted territory for all of us." In 2013, the German chancellor received much criticism for saying this sentence, as observers saw it as a reflection of how far behind the country is when it comes to digitalization. AI is the next digital technological frontier, one many feel will transform the entire economy,[16] and Germany is again not perceived as a global technology leader.

Key headline indicators are underscoring who leads in different fields: Total private AI investment in Germany corresponds only to a fraction of US or Chinese volumes (2.5% and 3.6%, respectively).[17] The list of top AI-patent applicants is dominated by Japanese and US companies. Only two German firms (Bosch and Siemens) made it onto the list.[18] While in the US and China, 35 unicorn start-ups in the field of AI have emerged since 2014, Germany is host to none – same as India.[c]

Many applications of AI that have already moved out of their piloting stage are in the B2C sector, drawing on behavioral data from e-commerce and other online services. In Germany, these sectors have been slow in responding to the "Web 2.0" phase and hence lacked even the data to be at the forefront of AI development. In car manufacturing, the US firm Tesla and the technology giant Alphabet received most of the attention for their autonomously driving vehicles – not the German original equipment manufacturers (OEMs).

In Germany, data economies á la China or the US are unlikely to materialize: While cross-border mobility of data is widely accepted in the EU and Germany, potential harms from excessive data sharing that goes beyond people's control need to be balanced. In this context, Germany will have to develop its own strategic approach to gain

c A unicorn start-up or unicorn company is a private company with a valuation over $1 billion. As of March 2020, there are more than 400 unicorns around the world. Variants include a decacorn, valued at over $10 billion, and a hectocorn, valued at over $100 billion. https://www.cbinsights.com/research-unicorn-companies

access to larger and more diverse data sets across industries – while at the same time adhering to its goal of developing responsible "AI made in Germany."

Against this background, this study explores the potential of such a bilateral exchange of data for AI between Germany and India.

2.4. Potential of German-Indian data exchange for AI

Before diving into assessing the potential of a data exchange partnership between Germany and India, it would be useful to understand why India represents a good case study for exploring such bilateral partnerships. Putting aside obvious aspects such as India's large population, democratic system and internationally successful IT-services sector (including the talent behind it), the country is a promising candidate for data cooperation with Germany for three reasons:

First, **India's digitalization efforts** have jump-started a vast domestic digital economy which also rests on an established digital baseline infrastructure called India Stack. This set of application programming interfaces (APIs) allows public sector actors, businesses, start-ups and developers to build systems to digitalize interactions, including identity verification, cashless payments and verification of documents. These steps contributed to the emergence of one of the most vibrant FinTech industries in the world, one with an estimated transaction volume of $66 billion.

Second, **India's political initiatives** have promoted progress in domestic AI-development (national AI-strategy, National Artificial Intelligence resource platform) while establishing rules for data protection and privacy in the country. The country has also launched several sector-specific policies (e.g. e-commerce, health) to clarify issues around cross-border transfers.

Third, **India's digital start-up landscape** has developed strongly over the past years, leading to the emergence of 21 unicorn start-ups, including several that collect large amounts of data potentially useful for AI, such as Ola Cabs and Ola Electric (mobility), Big Basket (e-commerce) and Pine Labs (FinTech). There are also new promising start-ups along the data value chain that are engaged in data labelling and curation. Although not the focus of this study, these companies are sources of significant expertise and understanding of data in India.

Taken together, these reasons and the aspects mentioned above position India as an interesting partner for increasing Germany's access to data for AI development. If realized, India's data potential – driven by end consumers (e-commerce, health), by its political priority setting for building a digital economy and by its IT-services industry – could prove to be a valuable part of Germany's own approach to AI leadership. A collaboration would provide a good fit with some of Germany's AI strengths (e.g. in health-tech) and might allow Germany to catch-up in sectors where the country has been lagging behind (e.g. e-commerce). Other areas could include mobility/traffic and agriculture. However, as the following sections show, the groundwork for such a collaboration with India must be laid before the potential can be materialized.

2.5. On the nature of a cross-border data cooperation of equals

Ever since data was first deemed the "oil" of the current economic era, an "extractive narrative" has dominated the debate: Data is valuable for those who mine it, a source of new products, and thus grants power to those who have it. From such a perspective, data represents a good to be extracted, hoarded and protected. Data analytics, machine learning and artificial intelligence are hence the technologies used to extract value from data as a raw material.

When speaking of India as a "data-rich" country – a statement that has motivated many actions in addition to the publication of this study – we always run the risk of following such a narrative. Moreover, a one-dimensional framing of India as a strategic reservoir of data for the development of AI in Germany or the EU comes dangerously close to echoing historical colonial practices of economic extraction. It is no coincidence that researchers and activists from various disciplines and regions have warned against an increasing "digital colonialism,"[19] "data colonialism"[20] and "algorithmic colonization."[21] And the extractive narrative is not just linked to relations between countries. A simple Global North-South dichotomy does not adequately describe a digital economy that tends to be dominated by relatively few globally active technology corporations whose headquarters are primarily in the US and China.[22] The extractive practices visible in many business models in the digital economy are also applied to those living in the "home countries."

Any meaningful Indo-German data partnership for AI will have to consider these different power dimensions in the digital world. Our study, which examines the baselines of such a collaboration, was written to help create a common knowledge base for a future, balanced partnership of equals. AI is a technology that can lead to discriminatory, marginalizing and outright violent outcomes. Both countries have subscribed to a vision of AI that reinforces ethical values and benefits all of society (most recently in co-launching the Global Partnership on AI in June 2020). This vision also needs to be embedded into the efforts meant to create a closer data exchange for AI development.

BOX 1 METHODOLOGY OF THE STUDY

This study has been conducted with a deliberate focus on the perspectives of data and AI practitioners. The goal was to understand the potential for data collaboration based on the regulatory landscape and actors present in the country.

We identified the Crunchbase dataset as a useful resource for analyzing the AI-related private sector landscape. The entire data set for India covers about 30,000 companies headquartered in India, with 2019 being the last year available.[23] The detailed description and categorization of each company's main focus provided us with an overview of the current data/AI situation.

In a second step, we interviewed experts working in and around the field of AI to understand the practitioner's perspective of using cross-border data to build AI. These interviews were conducted on the phone or in person and they lasted from 20 minutes to an hour. People with a technical or business background working in data science or AI were asked questions related to the kind of data sources used to build AI applications. We also spoke about major actors in the data value chain, as well as examples of data exchanges or cross-border collaborations to build AI applications. From a technical perspective, we explored ways to overcome roadblocks that prevented data from other countries from being used for building AI. A smaller subset of experts was interviewed about the policy and regulatory landscape.

3. Requirements for cross-border data sharing for AI

Over the past decade, digital ecosystems and digital governance have fundamentally changed. Before looking into the Indian regulatory and company setting (section 4), we outline findings about regulatory and ethical requirements for cross-border data transfers from the European/German perspective (box 2 outlines technical data requirements for AI).

3.1. Data regulation with regards to cross-border transfers of data

3.1.1. Personal data protection

The past few years have seen a strong increase in countries creating data privacy frameworks which aim at governing the control and use of (personal) data. More and more countries are regulating data flows within and beyond their borders, with a notable trend towards approaches relying on adequacy (i.e. "a public or private sector [body] finding that the standards of privacy protection in the receiving country are adequate" and that data can therefore be transferred).[24] Such regulation can affect a wide range of data-sharing practices which might be useful for developing AI algorithms. However, the overall economic effect depends on the de facto implementation.

The General Data Protection Regulation (GDPR) has altered the perspective on cross-border transfers of personal data:[25] The EU-wide framework has not only harmonized data protection for one of the largest economic regions of the world, it also has global reach, as it "applies directly to cross-border commercial transactions involving personal data from the EU, even if an organization operates from outside the EU."[26] Implicitly, the regulation also carries values and principles from the EU to other countries around the world.[d]

The GDPR continues the trend towards linking cross-border data transfers to conditions. We have identified three possible ways that cross-border data transfers can be considered lawful: The data is transferred to an "Adequate Jurisdiction," the data exporter has implemented a lawful data transfer mechanism, or an exemption applies.

Adequate Jurisdiction: The European Commission may decide that the transfer of personal data between the EU to a specific third country is allowed in a general manner, because the country in question has a level of data protection that is essentially equivalent to that guaranteed by the EU.[27] The factors influencing this decision include adequacy of the rule of law and legal protections for human rights and fundamental freedoms; the extent of access to transferred data by public authorities; the existence and effective functioning of data protection authorities (DPAs); and international commitments and other obligations in relation to the protection of personal data.[28] Major countries that have

d The following statement by the EDPS Ethics Advisory Group from 2016 illustrates this: "This new data protection ecosystem stems from the strong roots of another kind of ecosystem: the European project itself, that of unifying the values drawn from a shared historical experience with a process of industrial, political, economic and social integration of States, in order to sustain peace, collaboration, social welfare and economic development." See: O'Hara, Kieron, and Wendy Hall. 2018. "Four Internets: The Geopolitics of Digital Governance." CIGI Papers 206 (December). https://www.cigionline.org/publications/four-internets-geopolitics-digital-governance.

been recognized as providing adequate protection include Argentina, Canada, Israel, Japan, New Zealand, Switzerland, and Uruguay (the Privacy Shield framework with the US was invalidated in July 2020 by the European Court of Justice). In 2009, even before the GDPR was introduced, India sought adequacy status, but the results have not been published.[29] As outlined in section 4.2, there are doubts whether the current draft of the Indian PDP bill would overcome this hurdle.

Lawful data transfer mechanism: For data transfers within a corporate group or a cross-border organization, national DPAs can approve permanent rules for transfers of data if certain requirements are met.[30] Once such binding corporate rules (BCRs) are approved, no further DPA approval is needed for transfers of personal data covered by these rules. [e]

Derogations or exemptions: Even if none of the above applies, data transfers might still be possible if the individual whose data is to be transferred provides explicit consent for the transfer. Other exemptions include contracts between the individual and the "data controller," data transfers where public interests are concerned or where legitimate interests of the data controller exist. Importantly, these exemptions are intended for specific situations.

3.1.2. Local storage of data

Another emerging element to regulation around data in the EU and Germany – and India – is the requirement to "physically" store and retain at least some part of the data in the EU. Such local storage requirements come in different forms and can vary across different types of data. The OECD outlines a spectrum of different regulatory regimes present around the globe, which each have different implications for cross-border flows:[31]

Local storage requirements, no restriction of flows: A copy of the targeted data is stored in domestic computing facilities, but the transferring or processing of copies of the data abroad is not restricted.

Local storage requirements, no restriction of flows for data processing: Foreign storage of the data is not permitted, but transferring the data temporarily for processing is allowed. Data must be returned to the home country for storage.

Local storage and processing requirements, with restriction of flows: Foreign storage is not permitted, and the permission to transfer data is subject to conditions.

The narratives behind local data storage policies vary by political actors, but they generally focus on national/regional competitiveness rather than on data protection for individuals. In the current political debate, policies focused on the local storage of data are linked to domestic cloud infrastructures, digital/data sovereignty and technological sovereignty. The EU's Strategy for Data highlights these aspects clearly by stating that "the EU needs to reduce its technological dependencies in these strategic infrastructures, at the centre of the data economy."[32] The proposal by Germany's Federal Ministry for Economic Affairs and Energy (BMWi) for a federated data infrastructure (GAIA-X) stated similar goals: striving for data sovereignty and reducing dependencies on (non-EU) providers of data infrastructure.[33]

As these policies are still being discussed, the implications for international data transfers are not yet clear. The practice of physically storing data on servers located within a country (data localization) has been criticized as a major obstacle to data exchanges and to digital innovation in general. International industry associations have already raised their concern about such policies.[34] For example, Claudia Biancotti from the Peterson Institute for International Economics argues that "it [data localization] cuts into the revenues of foreign corporations, chiefly by preventing them

e Other lawful data transfer mechanisms include Model Contract Clauses, which do not need separate authorization from DPAs. EU Model Contract Clauses are like an addendum to contracts between European Economic Area (EEA) and non-EEA entities. They are designed to provide appropriate safeguards for the protection of personal data.

from pooling data from multiple countries when training artificial intelligence algorithms."[35] Less pointedly, but with more facts, the EU Commission Staff Working Document complementing the Communication on "Building a European Data Economy" from 2017 summarizes evidence on the economic impact of data localization requirements.[36] Much of this evidence focusses on the impact on cloud services. While these services are an important element of the current data economy, it also needs to be considered that these markets are extremely concentrated. For example, the global Infrastructure-as-a-Service market (e.g. public cloud services) is dominated by four actors who control 75% of the global market.[37] From the perspective of cross-border data collaborations, this situation results in uncertainty among firms and organizations.

3.1.3. Trade agreements

The introduction of the GDPR has had a profound impact on trade agreements by drawing attention to potential conflicts between the current realities of the digital economy and privacy standards. This situation has created some disagreement during ongoing trade negotiations, where some observers perceive a general conflict between the US preference for free flows of data without localization requirements and the EU's privacy regulation binding the adequacy decision to local conditions.[38] As these trade agreements are still being debated, it remains to be seen what stance the EU will take on India.

3.2. Ethics of data and AI

There is an emerging debate about the ethics of data and AI, as it has become increasingly obvious that data protection measures might not be sufficient to cover all relevant aspects when societies undergo digital transformation. It would exceed the scope of this paper to discuss the subject in detail, and treating it in outline would not do justice to its importance. Thus, we refer the reader to the research and debates that have been ongoing on different levels:

A recent article identified 84 documents containing ethical principles or guidelines for AI, 88% of which were released after 2016.[39] On the European level, the European Commission's "Ethics Guidelines for Trustworthy AI" were published in 2019.[40] In 2020, the Bertelsmann Stiftung and VDE, a nonprofit standards-setting organization, have been engaging in an interdisciplinary effort to move the discussion around the guidelines forward, so the principles can be operationalized in AI.[41]

Both Germany and the EU have declared that they want to develop and use "AI for good and for all."[42] As for this study, the topic will also become relevant with regards to cross-border data transfers with India. The integrity of the goals set by Germany and the EU will have to be compared to the de facto reality of how data is collected, used and protected in India. A semantic analysis of both countries' ethics statements in their AI strategies can be found in section 4.

3.3. Companies' willingness to trade and share data

Data sharing between businesses is one of the key drivers for AI developments. According to a McKinsey report from June 2019, 18% of large companies in the EU are using AI at scale in at least one function.[43]

However, in a study for the European Commission which surveyed 100 business organizations, the authors reported that only 11% of the companies were trading data in business-to-business relations, and 2% had adopted an open data policy to share data.[44] A mapping of data-sharing initiatives in Germany found no dominant player serving as data marketplace among the 13 identified data marketplaces.[45]

These low numbers aside, there are emerging data-sharing models that might allow the sharing of data sets meaningful to AI, namely when companies agree to strategic and collaborative partnerships in which data is shared

in a closed, exclusive and secure environment. One example from Germany is the strategic alliance between the car parts producer Dürr, the equipment manufacturer DMG Mori, the software development firm Software AG, the electronics company ASM PT and the measuring device company Zeiss. Together they have been building an open IoT platform that allows data sharing between equipment manufacturers, their suppliers and their clients in order to move towards digitally connected production.[46] Given how high the requirements for data are when developing AI models, these targeted approaches focused on one domain appear to be a productive way forward.

BOX 2 TECHNICAL REQUIREMENTS OF DATA FOR AI

The current workhorse technology within the realm of AI is machine learning – particularly neural networks. There are countless examples from a variety of industries that have proven that the current AI applications can be usefully deployed.[47] For the moment, machine-learning applications are essentially prediction algorithms where "prediction" is often not so much about saying something about the future, but rather about filling in "missing data" which can be derived from the training data. The translation of an English website into Hindi is such an example: The algorithm fills in the "missing" text in Hindi based on texts that have already been translated in both languages.

Such "prediction machines" have become a new commodity.[48] Multiple actors in the AI field have put out trained models (e.g. Google's AutoML Vision, an API that identifies or classifies objects in pictures) which can then be used by analysts with smaller data sets so they can adjust their algorithms to the problem at hand.

Despite the progress made in the underlying technology, major challenges persist that limit an even more widespread expansion of its use:

Large datasets: Current machine-learning methods are usually very data hungry. For example, a deep neural network in India needed a training data set of 1.15 million chest X-ray scans and the corresponding radiology reports. In the end, the algorithm was at least as accurate as four radiologists (standard of reference) for the interpretation of four different key findings for chest X-rays.[49] But the figures illustrate the expansiveness of the required data.

Accurate labels: Labelling such vast datasets in an accurate way can be a significant challenge in itself. In the health domain, for example, labelling the data often requires qualified specialists (in contrast to efforts to recognize animals in pictures). Thus, significant investment is often needed when no labelled data set

exists. For businesses, this increases the incentive to ensure that the returns on such an investment remain with the firm by excluding others from using the data set.

Legacy systems and comprehensive data sets: Quantity is not everything. Not all data collected, labelled and processed is suitable for the development of AI algorithms. In fact, only a fraction of the available data sets are comprehensive enough to allow the development of a useful application. A generic challenge from manufacturing illustrates the challenge of patchy data sets: Detecting defective parts produced by an automated process is a constant goal of quality control teams. During the regular production process, massive amounts of data are generated every day, such as visual data or data on pressure and temperature. This wealth of information (if it is even stored at all) is often not generated in comparable environments: Machines along the assembly line break down, workers' performance and inputs vary (un)systematically, or slight variations in the product require a change of input components. Every production process is subject to constant alterations. Theoretically, all these changes could be modelled into an algorithm. The reality in manufacturing today, however, is that structural breaks in the data set are very frequent. Production processes are very rarely designed with the intention of feeding data into a prediction algorithm, and retrofitting existing processes often fails due to limitations in the data describing the production process and the essential environmental parameters.

Generalizability of learning: Many of the current AI applications have emerged from data sets that were highly context-specific. Whether a data set that proved useful in training a model in one context will remain useful in another is questionable. For example, the accuracy of an image recognition algorithm that can identify number plates of Indian cars and two-wheelers is likely to drop dramatically when used in a European setting.

This remains a major question when assessing the potential of cross-border data exchange between Germany and India. Which data sets have value for AI independent of the context in which they were generated? In most cases, this question can only be answered after a test has been conducted in the relevant real-world conditions. While several firms and organizations have built quite substantial lists of open data sets on a wide variety of issues,[50] the question of validation in the local market still persists. As Rahul Alex Panicker from Wadhwani Institute for Artificial Intelligence in Mumbai argues: "Even if you want to adopt solutions that are being used elsewhere, you want to make sure that they continue to be valid in your context and, for that, you also need local data sets."[51]

Risk of bias: Training models based on data from a different context comes with its own risks. This is particularly important for behavioral data generated by users: Consider a machine-learning model that uses a text data set generated in India to build a chatbot that interacts with people who wish to execute a financial transaction through their bank. The model is able to identify certain questions and predict the likely answer. Using such a data set to train a model for the German context might introduce a systematic bias into the prediction, because the original data was predominantly generated by young Indian men from a relatively wealthy layer of Indian society. In Germany, the requests to the chatbot might come from a different group of people who lie outside the algorithm's "experience" and hence receive systematically worse answers. Such risk – if not considered – will significantly damage the trust in the application. How was the data collected? With what goal? Who is part of the data set – and who has been left out? Answers to these questions determine whether the data introduces bias into the automated process.[52]

4. India's digital momentum

This section provides in-depth analyses of the developments in India with regards to the country's digitalization efforts (4.1) and major policy and regulatory decisions (4.2). The situation in the AI-related tech companies is described in section 4.3.

4.1. India's data potential from consumer-led digital transformation

Since the 2000s, India has undergone rapid digitization. Three major efforts give an idea of the scale at which this is happening: First, over 370 million bank accounts have been opened since 2014 as part of the financial inclusion program Jan Dhan Ayojana. Second, the unique identity-number program Aadhaar, which was started over 10 years ago, now covers over 1.25 billion people in India.[53, 54] Third, there are almost 1.2 billion mobile phone subscriptions in the country and over a third of the Indian population uses the internet.[55]

From the government side, this "trinity" – higher penetration of bank accounts among previously "unbanked" people, greater digital identity, and more mobile phone use – opened up the opportunity for technology-enabled direct benefit transfers (DBTs), with the goal of curbing corruption when disbursing subsidies.[56]

A nationwide digital baseline infrastructure (India Stack) was started in 2009 with the Aadhaar "Universal ID numbers." It represents a set of APIs allowing governments, businesses, start-ups and developers to build on four essential systems to digitalize interactions: an identity verification system (e-KYC), a system for legally binding electronic signatures (Esign India), a system for cashless payments (UPI) and a system for verification and issuance of documents and certificates (Digilocker).[57]

The financial sector is probably one of the most impressive examples of the success of these systems: Traditionally, banking in India has had limited penetration because of the lack of identity verification systems. Banks were custodians of the "identity" of those who had bank accounts. This gave banks an edge when it came to upselling and cross-selling other products, such as financial insurance or investment. India Stack provided a solid starting point for the development of a vibrant FinTech-sector, as it allows citizens to open a bank account or brokerage account or buy a mutual fund anywhere in India with just a fingerprint or retinal scan from Aadhaar, which also provides e-KYC services.[58, 59]

This created a more level playing field between established banks and newer FinTech firms enabling greater competition between the two. From January 2013 to October 2018, approximately 2,000 FinTech companies were founded.[60] According to a 2019 report by consulting firm Ernst and Young, India ranks second globally when it comes to adoption of FinTech services – almost on par with China.[61] It is estimated that the transaction value in the Indian FinTech market is about $66 billion (estimates for Germany are around $120 billion).[62] The CEO of Alphabet, speaking of Google Pay in India and thereby Unified Payment Interface (which is a

part of the India Stack ecosystem), said it is a globally relevant model that the company wishes to replicate in the US.[63]

There were further developments that have boosted the digital economy: the increasing average income (particularly in urban areas),[64] the dramatically falling price of internet data packages (more than 95% since 2013),[65] and 4G coverage of more than 90% of the total population,[66] all of which create the opportunity for a consumer-driven digital transformation. Cash payments halved from 59% in 2000 to 30% in 2016.[67] Mid-2018, the number of transactions in e-commerce retail was 1–1.2 million per day. There were 55–60 million transactions per month on e-commerce platforms.[68]

Analysts have projected that the Indian digital economy (including telecommunications and electronics manufacturing) could grow to $355–435 billion by 2025, driven not least by the IT sector and the "business process management" sector.[69, 70]

India is also pushing other data-heavy areas such as genomic research: Last year, the Indian government announced the financing of the IndiGen project led by the Council of Scientific and Industrial Research (CSIR), which concluded sequencing of whole genomes of 1,008 people by the end of 2019 and intends to sequence at least 10,000 genomes over the next three years.[71]

Other areas have remained rather stagnant: The manufacturing sector has seen little change in its outlook. While "Make in India" was a flagship project initiated by the first Modi government and the manufacturing sector has been growing, manufacturing has not become more important for the economy as a whole: Between 1991 and 2018, the share of the manufacturing sector in the country's GDP stagnated at 16%.[72] Another indicator by the research firm Economist Intelligence Unit expands the picture by examining the innovation environment in automation: While Germany ranks third after Japan and South Korea, India ranks 17th out of 25 selected economies – behind Turkey and Malaysia.[73] A report by the Bertelsmann Stiftung on the "Future of Indo-German Industry Collaboration" from 2018 summarized the data-related concerns as follows: "Indian companies face the problem of capturing data at machine level. […] Most stakeholders agree that data is at the core of Industry 4.0, be it data analytics, machine intelligence or deep learning but there is not enough focus in companies on these topics."[74]

In summary, India's increased data potential has its roots in the inclusion of a larger share of its population in the digital economy. The sheer size of India's population and rapidly growing intensity with which it has adopted digital services also offers potential for accessing data relevant to AI development.[f]

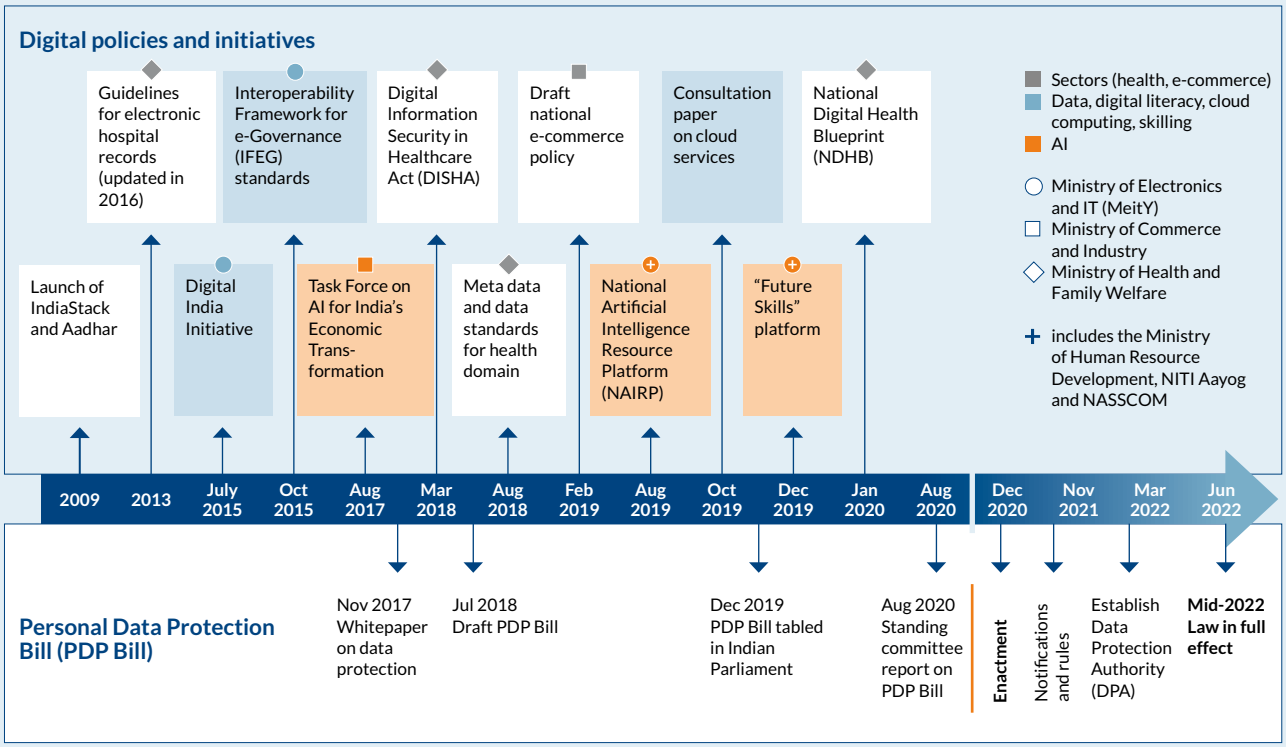
4.2. India's data and AI policies

This section gives an overview of the major developments for data- and AI-related policies and regulation in India. Just as in many other countries, our analysis shows that the policy and regulatory environment in India has not developed as rapidly as its digital transformation. This also has immediate effects on the potential for cross-border data exchange for AI between Germany and India.

The timeline below shows milestones towards the enactment of the major policy in this regard, the PDP bill, as well as other regulatory and policy projects. Key actors in the discussion are the Ministry of Electronics and Information Technology (digitalization and AI initiatives), the Ministry of Commerce and Industry (e-commerce and AI), and the Ministry of Health and Family Welfare (e-health). Further actors included here are the government think tank NITI Aayog and the powerful software and IT-services industry association NASSCOM.

f This is reflected in the selection and content of our data deep dives (see below), where we selected retail and health as two sectors heavy in consumer data.

FIGURE 1 **TIMELINE OF MAJOR POLICIES AND REGULATORY MILESTONES**



Source: Research by CPC Analytics.

BertelsmannStiftung

Data privacy and protection

On December 11, 2019, the Personal Data Protection (PDP) bill was first presented to the Indian parliament. It has now been referred to a joint parliamentary committee for further debate and examination. The committee has been instructed to give its report to the Lok Sabha for the Budget Session 2020. This follows an August 2017 ruling by the Indian Supreme Court which held the right to privacy to be a fundamental right under the Constitution of India. In a follow-up process, the Sri Krishnan Committee Report on data protection established the basis for the current draft.

Concerns raised by several companies and global trade bodies have been submitted to a joint parliamentary committee that will present its report in the monsoon session (June 2020 onward) of the Indian parliament. Following this, if the bill receives a majority vote in parliament, it will become an act and receive presidential assent. The notifications and rules of the PDP Act would be drafted by December 2021 at the latest and of the data protection authority (DPA) by March 2022. The law can be expected to come into full effect between March and June 2022.[75]

The following analysis focuses on key elements of the PDP bill that might have an impact on cross-border movements, while contrasting these elements with the GDPR.[76]

- Territorial and material scope: Like the European GDPR, the Indian PDP bill would imply extra-India application, i.e. it would apply to all entities that have business connections to India or conduct data profiling on individuals in India. The PDP bill covers

personal data and non-personal data to a certain extent, while the GDPR does not cover non-personal data at all.[g]

- Cross-border data flows and local storage of data: In general, the PDP bill does not place any restrictions on the transfer of personal data per se. However, it establishes two types of personal data which come with transfer restrictions. Sensitive personal data may reveal, be related to or constitute health data, financial data, genetic data, biometric data, sexual orientation, etc. In this case, the data needs to be stored in India, but may be transferred outside for processing if the individual has given explicit consent. Critical personal data does not exist in the GDPR and is not further defined – except for the statement that it means "such personal data as may be notified by the Central Government to be the critical personal data."[77, 78] Such data may only be transferred outside India in case of an emergency affecting the individual or following a decision by the government. As with the GDPR, the data protection authority (DPA) can either approve an intra-group program for transfer, or the government must approve the transferring entity or country. A key difference to the GDPR is that the rules for "adequate jurisdictions" are less clearly defined in the PDP bill.

- Innovation sandboxes: The PDP bill does foresee a mechanism that would allow entities controlling personal data ("data fiduciaries," which corresponds to "data controller" in the GDPR) to have some obligations relaxed for a period of up to three years.

In its current state, there are three major concerns about the PDP bill: (1) data localization, (2) government access to non-personal data and (3) the classification of sensitive and critical data.

(1) The localization clause is contested given its proposed strict application that could hinder the free flow of data across borders. Other countries would recognize data transfers as the norm and only apply restrictions to "either address inadequacies in the legal regime or in the privacy practices of the recipient." On the other hand, in the PDP bill, localization is the default, requiring data handlers to store a copy of sensitive personal data in India; it also restricts the movement of data across national borders in most cases. This could significantly increase the cost of compliance for foreign players.[79, 80]

(2) The second objection pertains to the clause requiring companies to share non-personal, anonymous data with the government of India. The bill does not specify the conditions under which the government can access personal and non-personal data, and this excessive power and ambiguous definition could result in privacy concerns for citizens and businesses.

(3) Finally, the category of "critical personal data" has been introduced, but its scope is not clearly defined. This critical data cannot be transferred under any conditions, and the lack of clarity on its definition creates uncertainty for businesses.[81]

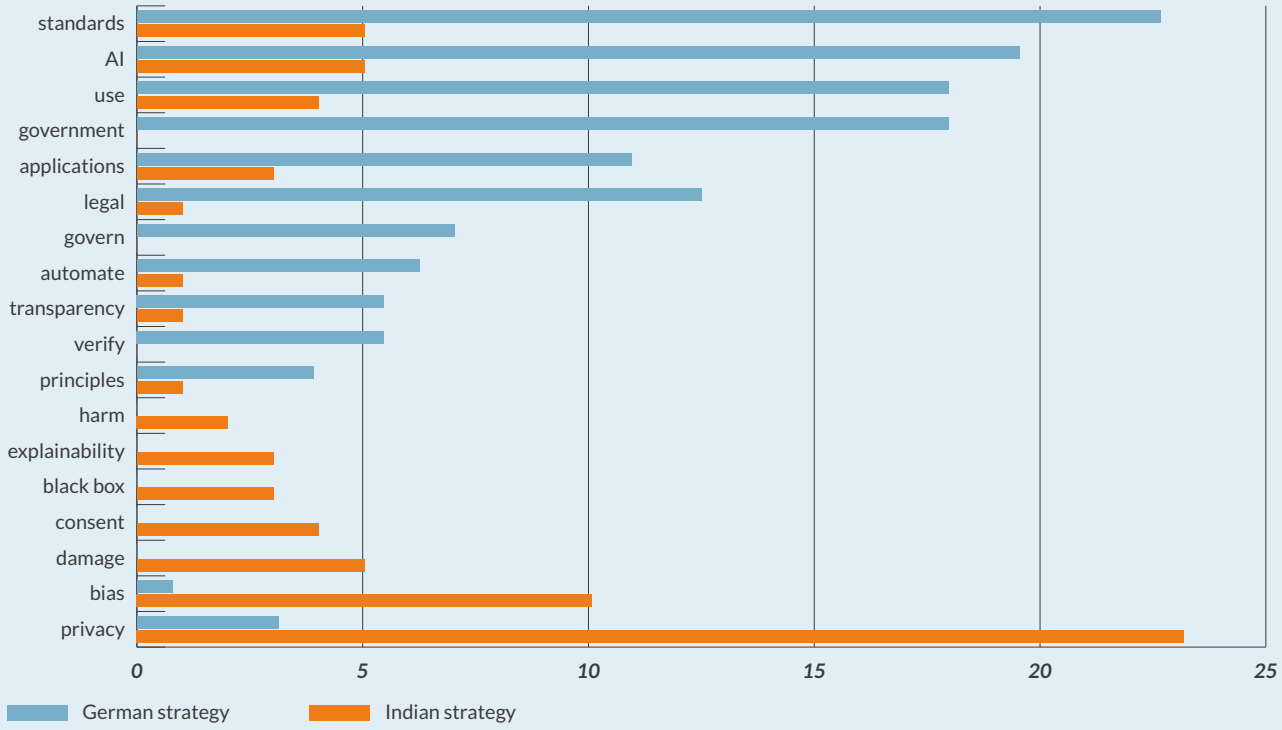
Particularly the latter two points represent a risk as to whether India would be recognized as an "adequate jurisdiction" by the European Commission.

Strategy fostering AI development in India:

In June 2018, the government think tank NITI Aayog presented a working paper titled "National Strategy for AI: #AIforALL" which is commonly referred to as "India's AI Strategy."[82] It is supported by a wide range of policy actors and is therefore expected to have significant effects. However, it does not state any budgetary figure to be invested in the field of AI.[83] Out of over 30 recommendations, the following three are of particular interest for cross-border data exchanges and international technology cooperation:

g See Section 91 of the 2019 PDP draft: Importantly, the section empowers the Central Government to request any data fiduciary (= "data controller") to provide any anonymized personal data or non-personal data "… to enable better targeting of delivery of services or formulation of evidence-based policies by the Central Government, in such manner as may be prescribed."

FIGURE 2 **DIFFERENT PRIORITIES IN THE "ETHICS OF AI" SECTION IN NATIONAL AI STRATEGIES**
RELATIVE FREQUENCY OF KEYWORDS PER 1,000 WORDS



Source: National strategies, analysis by CPC Analytics.

BertelsmannStiftung

- Focus areas: The strategy identifies five priority sectors: healthcare (with a special focus on access and quality), agriculture (farm productivity and waste reduction), education (access and quality), smart cities and infrastructure, and mobility and transportation.

- Research institutions: By establishing Centers of Research Excellence (COREs) and International Centers for Transformational Artificial Intelligence (ICTAIs), the Indian government aims to boost fundamental research (COREs) and create applications with societal importance.

- National AI Marketplace: COREs and ICTAIs are to set up a "consortium of Ethics Councils" to develop sector-specific guidelines on privacy, security and ethics leading to a National AI Marketplace (NAIM). A "three-pronged, formal marketplace could be created focusing on data collection and aggregation, data annotation and deployable models."[84]

Comparing the sections on ethics in both the German and the Indian AI strategies using a simple semantic analysis based on word frequency, interesting differences can be seen in how the topic has been addressed. While the Indian AI strategy puts a strong focus on privacy and (algorithmic) bias, the German strategy concentrates more on standards, governance, principles and legal aspects. Although the strategy is only a snapshot of what the Indian government is considering when it talks about ethics and AI, it is still a starting point for discussions between the two countries (see also box 3 for more considerations on ethics).

BOX 3 CURRENT STATE OF DATA PROTECTION IN INDIA

While designed for data protection, the GDPR takes ethical principles into account. For instance, it specifies the "fundamental right of data protection" in several detailed descriptions of the rights individuals have.[85, 86] Within the EU, Germany has been a stalwart upholder of data protection principles. Considering the ethical aspects of data trade/exchange with India is therefore a matter of integrity for Germany, as it wants to ensure that its activities do not conflict with the relevant principles or compromise the privacy of people in India.

India is in a phase of transition with regards to its privacy and data protection principles and laws. A landmark judgement by the country's Supreme Court in August 2017 recognized the right to privacy as a fundamental right under Article 21 of the Constitution as a part of the right to "life" and "personal liberty." "Informational privacy" has been recognized as a facet of the right to privacy, where privacy protection extends to information about a person and the right to access that information.[87] Thus, India has effectively enshrined privacy as a fundamental right, something that has far-reaching positive consequences for its future activities in the area of data privacy and the secure exchange of data. As mentioned, in 2017, India drafted a Personal Data Protection (PDP) bill, which is currently being discussed in parliament.

This is not sufficient, however, to conclude that the data collected in India and crossing Indian borders adheres to the expected ethical and data protection standards. For this, we need to examine the laws currently governing data in India:

India does not yet have a dedicated data protection law, and data in the country is protected by the Information Technology Act from the year 2000 and by the IT Rules from 2011, which are particularly relevant for data protection and cross-border transfers.[88]

The IT Rules of 2011 in combination with the IT Act lay down that corporations must possess a comprehensive privacy policy for handling personal information (Rule 4) and obtain the provider's consent before collecting personal information for a purpose connected with its own functions (Rule 5).[89] For disclosure of information to a third party,[90] the permission of the provider must be in the contract itself. Otherwise the third party cannot disclose this information further. Section 7 gives adequacy provisions for cross-border movement of data.[91] Section 43A ensures reasonable data safety procedures by providing for compensation in case of a failure to protect data.[92] Section 72A[93] applies to contraventions committed in and outside[94] India irrespective of nationality and penalizes disclosure of personal information without consent.

However, several experts have noted that the IT Act (drafted in 2000 and amended in 2011) is incapable of dealing with challenges posed by modern data analytics and AI. The act deals with privacy and data protection in a piecemeal way (the reason why the country is drafting its PDP bill). Some of the major points of criticism include:

- The definitions in the IT Act are not specific and comprehensive enough. Crucial terms like "consent" and "explicit consent" remain vague and could be misinterpreted.[95]
- Several technological aspects of digital life today are inadequately covered by the IT Act. For instance, issues such as cookie consent have not yet been addressed by Indian legislation.
- The IT Act applies to bodies and corporates only located within India,[96] which is a significant limitation.
- The government need not adhere to the act.[97] Government agencies can ask for data without consent, provided it is for the purposes specified. This is also a significant criticism raised against the PDP bill, as mentioned in the previous section.

In addition to these legal gaps, our interviews with experts revealed the presence of significant enforcement gaps. For example, an AI expert at an Indian health firm described the practice of receiving non-anonymized patient data. Asked for the reason for this practice, he indicated that the hospital did not have the means to anonymize the data. This example is indicative of multiple instances where lax and relaxed practices with regards to data privacy were identified.

Moreover, India has also experienced several data leaks and breaches. Despite the Aadhaar (Data Security) Regulations, under the 2016 Aadhar Act, journalists in India exposed a major breach in the national data system, allowing them to access personal data that had been submitted to the government by private individuals for a mere 500 rupees (ca. $8).[98] In February 2020, it was found that over 1 million patient records and medical images of Indian patients had been leaked online from major hospitals in India.[99] Several other such examples of leaks and breaches have exposed the frailty of India's data protection systems.[h]

Given this evidence, it seems clear that neither the current legal base for data protection in India (IT Act) nor the data privacy practices observed among stakeholders provide comprehensive protection for personal data. While strong implementation of the PDP bill and establishment of a DPA promise considerably higher levels of data protection in the country, the decision to promote a data partnership for AI between Germany/EU and India needs to consider these realities.

h These include the leak of 1.3 million credit card details on the dark web, location data for people using the dating app Grindr, and user information from Facebook. https://www.deccanherald.com/national/a-look-at-data-breaches-cyberattacks-india-saw-in-2019-785987.html.

4.3. AI-relevant company landscape in India

The following section provides an overview of the diverse company landscape which goes far beyond the IT services industry that has contributed much to India's visibility in global business affairs.

Three groups of companies emerged from our analysis (see figure 3). Only one of these groups is purely focused on AI – we labelled them "core-AI firms" because their product offering is all about applying AI/ML methods. A much larger group of companies ("sectoral AI firms") experiments with AI/ML in order to support its existing business. These companies started collecting data as a side effort to their regular activities and view AI/ML as an additional technology they can use. The third group of companies largely represents the aforementioned IT services companies in India, which are expanding their services into AI/ML as their international clients move into the technology.

4.3.1. Core-AI firms

This group of companies has machine-learning technology at the core of its product offering. Usually, these companies are highly specialized and combine deep domain knowledge with technical expertise. For example:

- Firms building automated technology for detecting lung diseases based on image recognition and selling this product to hospitals.
- Conversational chatbot firms that offer a software to banks to handle regular service operations with clients.
- Firms offering speech recognition applications transferring speech into text based on neural networks.

FIGURE 3 **CONCEPTUALIZATION OF THE AI-RELEVANT PRIVATE SECTOR**

AI-related firms

Core-AI firms	• Machine-learning technology is at the core of the firms' products. • Highly specialized, often combine deep domain knowledge with technical expertise. • Data is often a vital starting point for these firms.
Sectoral AI firms	• Machine-learning applications make their regular business activities more efficient. • Sometimes they expand on their offerings using ML/AI. • Data is often a side product of their regular business activities.
AI-as-a-service	• IT-services firms that offer a wide range of services. • Building machine-learning algorithms is one offering. • Data is mostly from their clients, not from within their own business.

Source: Analysis by CPC Analytics.

BertelsmannStiftung

We identified 890 core-AI firms in India.[i] Geographically, they are clustered in the country's major technological and economic hubs of Bangalore (30%), Mumbai and Pune (19%), Delhi and surroundings (16%), Hyderabad (8%), and Chennai (5%). They are mostly young firms and start-ups: 87% of "core-AI firms" were founded after 2011.

There is a significant difference in the sectoral distribution among these firms in India. We quantify this difference with the sector-wise density of AI companies compared to all companies in the sector (see figure 4). A higher value indicates that there is a higher number of core-AI firms in the sector relative to all companies in the respective sector.[100]

Unsurprisingly, the banking and finance sector ranks very high, as does internet services (e.g. cloud or IoT services). To give some illustrative examples of such companies: ZestMoney is a Bangalore-based consumer lending FinTech start-up. The company's offering is to use mobile technology and AI to provide people with short-term credit on purchases of specific products. In the field of internet services, haptik is a good example: The company builds chatbots – an AI-based product that did not exist a decade ago. In the health and biotech field, eKincare, launched in 2014, created an AI-powered "personal health assistant" which captures users' medical data, such as physical records and health history, and suggests steps for having a healthier lifestyle. The company's business model targets large companies that want to support their employees' health.

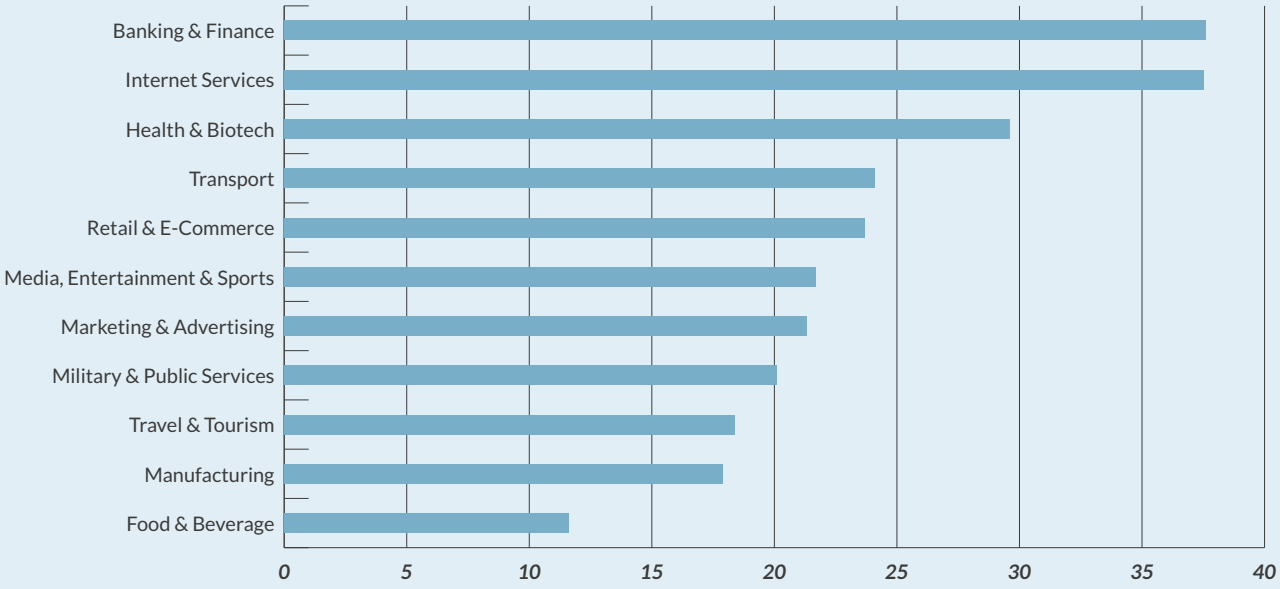
4.3.2. Sectoral AI companies

Many companies using AI and machine-learning applications make their traditional business activities more efficient or expand on their existing product portfolio by using AI. Data is often a by-product of every activity they engage in, but only now have they started collecting and curating it in a systematic manner so they can improve their regular business through multiple verticals.

A 2019 global survey of 3,000 CIOs found that 37% of companies have implemented some form of AI.[101] Another study found that corporate engagement with AI has shifted from "if" to "how" and the drivers of these projects have changed from the C-Suite to the IT department,[102] indicating that AI

i All these firms describe their products or services as AI/ML-powered or AI/ML-based. This definition is quite narrow compared to wider categories such as big data, analytics, etc. Data on revenue, capital structure, and employees is not consistent enough among the 890 companies to give a reliable picture.

FIGURE 4 **SECTOR-WISE DENSITY OF CORE-AI FIRMS IN INDIA**
AI FIRMS PER 1,000 FIRMS IN THE RESPECTIVE SECTOR



Source: Analysis by CPC Analytics based on Crunchbase data.

BertelsmannStiftung

development is being mainstreamed within some companies. Our interviews with sectoral experts in India indicate the same: Several sectoral firms either use AI or consider it a crucial addition to their existing business. Some examples include:

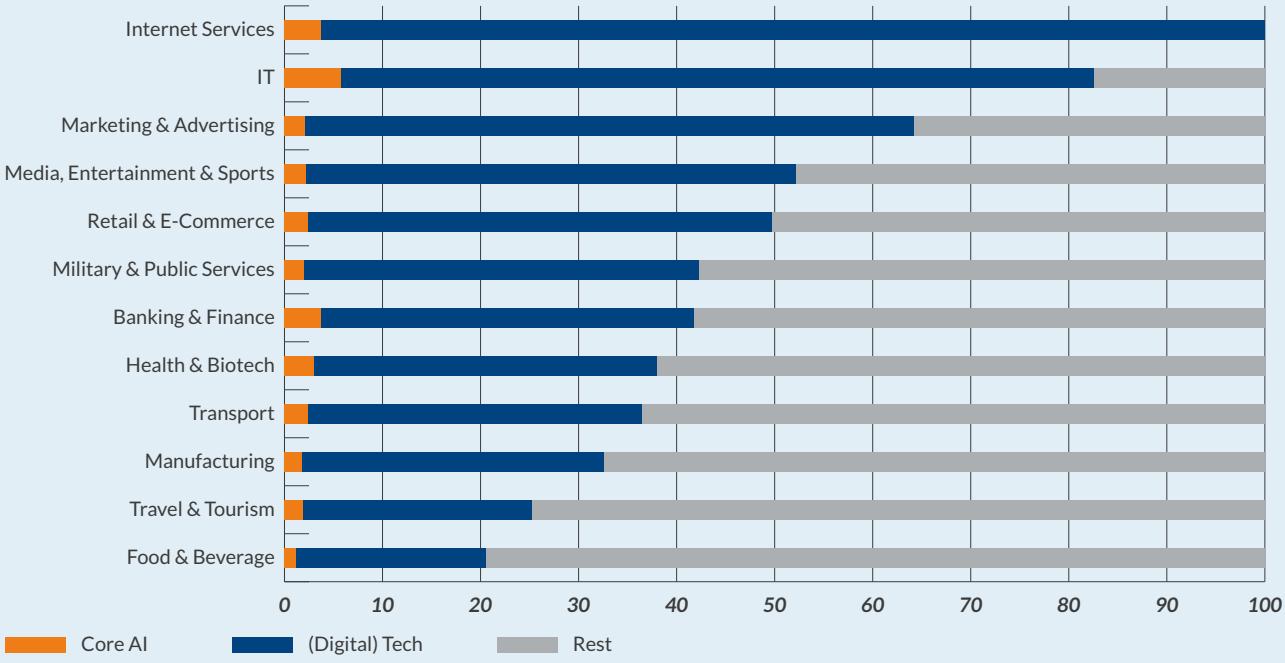
- Retail firms that improve the search function of their online shop using prediction algorithms that draw on previous shopping behavior.
- Equipment manufacturing firms that use machine learning to reduce the set-up costs of their machines once they have been deployed at the client's plant.
- Firms that offer IoT solutions connecting energy consumption monitoring across devices and use machine-learning models to improve their energy-saving recommendations.

Quantifying the share of companies that have the potential to build AI/ML based on their current business operations is inherently tricky. However, while every company collects some data simply by running their business, we are particularly interested in those that go beyond this – i.e. companies that use some variant of (digital) technology. The assumption here is that these companies will most likely be able to obtain sufficient data for building AI/ML in the near future.

Figure 5 gives the sectoral breakdown of such companies, indicating digital technologies as an important part of their product offering, next to core-AI companies, and the residual group of companies. By definition, companies offering services via the internet are digital. Hence, all companies in this sector potentially collect data. More interesting are the other sectors: A large share of companies in marketing (62%), media (50%) and retail (47%) sectors have digital technologies as part of their product offering. This is where we see the largest potential for big data from recording user/click behavior.

Health, banking, transport and manufacturing have also tremendous "data potential." However,

FIGURE 5 **SECTOR-WISE USE OF AI AND DIGITAL TECHNOLOGY IN INDIA**
PERCENTAGE OF CORE-AI FIRMS AND FIRMS THAT USE DIGITAL TECHNOLOGY, BY SECTOR



Source: Analysis by CPC Analytics based on Crunchbase data.

BertelsmannStiftung

the majority of the companies in these sectors do not traditionally rely on digital technologies. The relatively low share of digital technology firms in the manufacturing sector in India might be due to the identification strategy employed in our analysis: Any digital technology used in the manufactured products is either at the production level or does not constitute a major element in the product offering (and is therefore not captured by our identification method).

However, from the Crunchbase data set, it cannot be concluded that firms that do not feature "artificial intelligence" in their description or work categories do not use it at all. It is very likely that several firms that use digital technology embed AI in their operations to increase efficiency even if it is not explicitly mentioned. For example, Swiggy and Zomato, the dominant food delivery platforms in India, do not mention AI on Crunchbase, but use it for a variety of functions, from reducing delivery wait times to deploying chatbots for consumer queries and cancellations.[103] Hence, the numbers above represent a lower boundary for the companies that are in a good position to build AI-relevant data.

4.3.3. AI development as a service

This group includes mostly IT service firms that offer a wide range of software development services – among which building machine-learning algorithms is one offering. Some examples include:

- Tata Consultancy Services (TCS, an IT service company in India) created an IoT platform to collect and analyze data for Cargotec (Finland-based cargo handling company). The goal was to provide algorithm-driven actions, business process automation, and advanced real-time analytics on equipment data to increase operational efficiency.[104]

- Mu Sigma (a data analytics services and consulting firm in India) has used AI to produce various accelerators/code blocks for exploratory data analysis, automate video analytics and natural language queries, create classification and forecasting models, and develop chatbots for their clients.

- Infosys (an IT service company in India) consulted for a New Zealand-based telecom company to ensure successful AI adoption in their service delivery.[105]

These examples highlight firms that offered IT or analytics services across industries and are now extending their business to machine learning offerings. This shift comes with the rising demand for such products across industries. A 2019 study that spoke to 400 experts globally found an increasing "buyer" versus "builders" divide in AI: 44% of the respondents preferred to buy AI solutions from a third party and 11% preferred to outsource custom solution development to a third party.[106] This preference to buy rather than build indicates a demand for AI service companies, and erstwhile IT service firms are well-placed to meet this need.

The Economic Times India Leadership Council's core group posited that the Indian IT service industry is poised to take advantage of the growth in demand for AI services and technologies.[107] The industry currently builds software for business process automation worldwide. Even if the industry focusses on price as a competitive advantage, there is already a commodification of AI visible in some sectors: Low-cost AI chatbots for customer services, for example, are being offered by more and more companies.

5. Deep dives into Indian sectors

As data collection is streamlined, AI applications are increasingly being integrated into the value chain. There are many sector-specific nuances to this integration of data and AI that often cannot be generalized across the entire economy. For instance, in a consumer-led sector like healthcare, the actors, types of data, controlling mechanisms and laws are very different from those in the manufacturing sector, which relies on machines and is driven by efficiency considerations.

Two deep dives elaborate on the analysis of the AI landscape in India provided in the previous sections. The health and retail sectors were selected for two main reasons: In both cases, the ever-growing share of the Indian population participating in digital life is creating a wealth of data (behavioral and biological). Second, both sectors "produce" data of which a large share is likely to be comparable across countries.

5.1. Deep dive I: Data for AI and India's healthcare sector

AI applications have the potential to transform the healthcare sector across its value chain – from efficient delivery of services for predicting, diagnosing and prescribing care to easing the patient's experience when receiving care. The numbers reflect the potential: The global market for AI in healthcare was valued at $1.4 billion in 2016 and is estimated to reach $22.7 billion by 2023.[108]

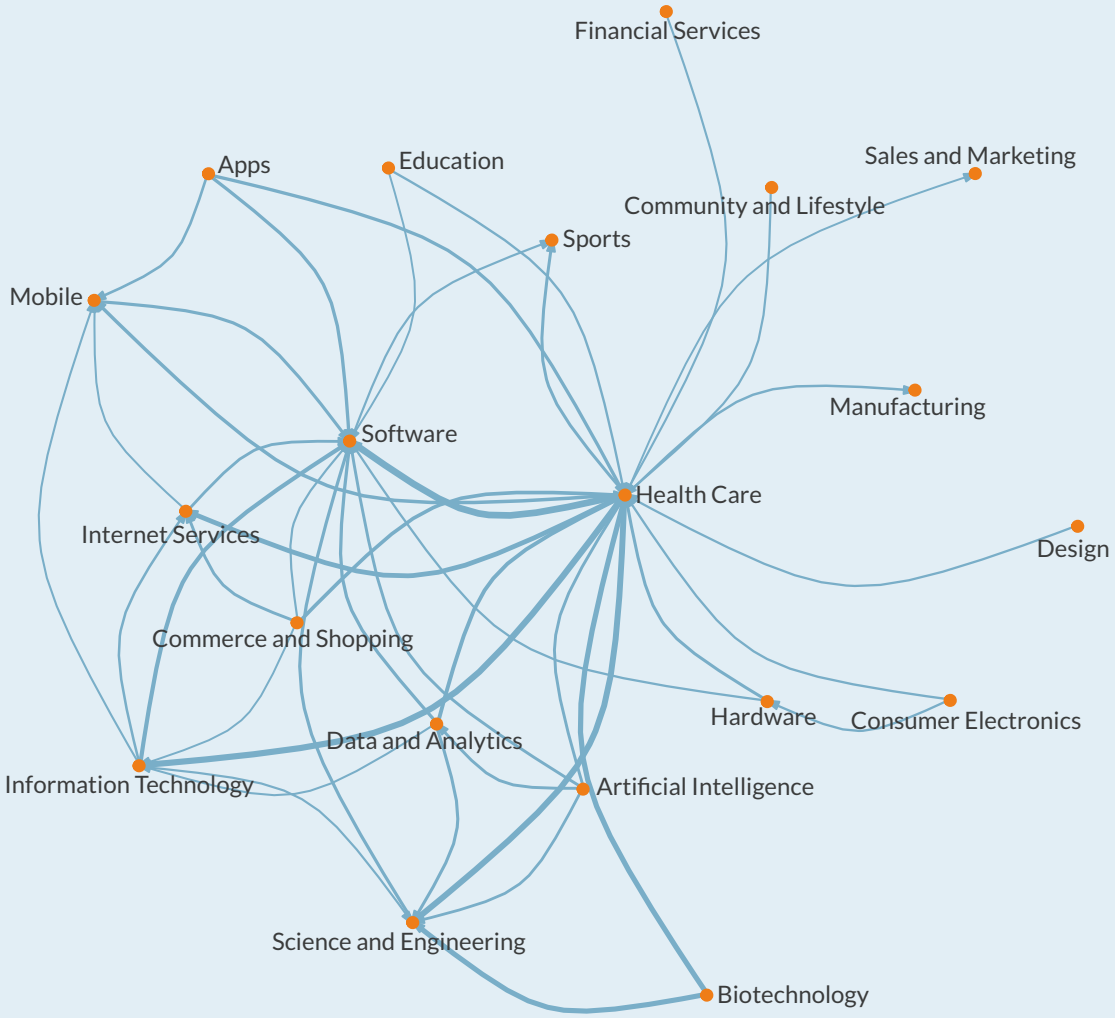
Structural aspects of the Indian healthcare system

With 1.3 billion citizens, India offers immense potential for using medical data to train and improve AI models. However, to assess the potential of working with data from India, we need to consider the current state of healthcare in the country. The Indian healthcare system is highly informal, suffers from a shortage of physicians and has a high out-of-pocket expenditure (69.3%[109]) and low government spending (1.4% of GDP[110]). There is also a huge network of private hospitals (70% of health spending is private), of which only 2% are accredited.[111] Health data in India suffers from the same issues that apply to data globally – albeit even more acutely. A vast majority of the Indian healthcare sector has poor-quality digitized hospital operations data (such as Hospital Information Systems, and Enterprise Resource Platform software) and patient data (Electronic Health Records). Digitization and "networked hospitals" are far from the current reality. This implies that the availability of machine-usable data is severely limited and, for most hospitals, incorporating AI in healthcare is a distant dream.[112]

Health-tech company landscape as an indicator of increasing AI adoption

While these structural challenges exist, a parallel economy of start-ups and health-tech firms is emerging in tier-1 and tier-2 cities in India. There are 53 one-million-plus cities with a total population of about 143 million.[113] Several private hospitals in these urban centers have made significant progress in digitizing hospital and

FIGURE 6 **SECTORS AND TECHNOLOGY THAT FIRMS IN HEALTH-TECH ENGAGE WITH**
The thickness of the line is indicative of the number of engagements between health-tech firms and other sectors and technologies. The top three sectors that health-tech engages with include software, IT and science and engineering.



Source: Analysis by CPC Analytics based on Crunchbase data

BertelsmannStiftung

patient data and, even though AI application is limited, they could provide a source of data for AI.[114]

Government forecasts estimate that by 2022 the healthcare market in India will grow to $372 billion, the medical device market will exceed $11 billion and the medical technology sector will reach $9.6 billion.[115] Other sources also estimate that the market for AI applications in healthcare is growing at around 40% and could reach $6.7 billion by 2021.[116] Medical devices, technology and AI would thus contribute to increasing the health-data pool in the country. The promise that Indian healthcare data holds is also reflected in private investment in the sector. In 2017, health-tech start-ups in India received total funding of $338 million, which was 2.5% of all start-up

funding that year. Of this, almost 31% of health-tech funding went towards artificial intelligence, IoT and analytics.[117] However, foreign direct investment (FDI) in hospitals and diagnostic centers has been relatively low, at just $6.63 billion for the 19-year period between April 2000 and December 2019.[118]

Since 2000, 66 health-related core-AI firms have been established in India. However, this start-up activity is only the visible surface of a larger development. Several major sectoral players like Manipal hospital group and Max Healthcare have also employed AI in their operations, even though it is not part of their core business. According to our analysis, there are 817 companies in India that specifically offer technology solutions in health. The network chart below provides an overview of companies in India working on digital health-tech, showing the main fields in which they are active.

Apart from Indian companies and start-ups, more MNCs and large ICT companies are now offering AI solutions for healthcare in India. Several companies have created tech products for the sector. For instance, Siemens Heathineers has used AI to develop a digital twin heart[119] and Philips sells AI-enabled heart models.[120] Several state governments in India have funded projects and are working with MNCs for diagnostics and preventive care. For example, Microsoft is working with the state government of Telangana on its Rashtriya Bal Swasthya Karyakram program, for which the state has adopted the AI-based Microsoft Intelligent Network for Eyecare (MINE) platform to reduce avoidable blindness.[121]

Data projects and sources

Even though the data infrastructure needed to support AI is limited and disconnected, some cross-border collaborations on focused problems or on a small and localized scale have yielded some promising results. This can be observed in examples in the private sector and at an intergovernmental level.

In the private sector, Google collaborated with three prominent eye hospitals in India (Aravind Eye Care, Shankar Nethralaya and Narayana Nethralaya) to develop its retinal screening system that detects diabetes-related eye disease. To do this, they used retinal scans from India to train their image-parsing algorithms and then used an open public data set from France to validate the results.[122, 123] This was possible by employing cross-border data and algorithm flows from entities in India, France and the US. Other examples include Microsoft's collaboration with Apollo Hospital (in India) to build an AI cardiology network, and IBM Watson's work with the Manipal Group of Hospitals (in India) to aid doctors in the diagnosis and treatment of seven types of cancer.[124]

At the intergovernmental level, the Pan-Cancer Analysis of Whole Genomes (PCAWG) project worked from 2013 to 2019 to create a database of 2,658 cancer genomes from 468 institutions and 34 countries in Asia, Australasia, Europe and North America.[125] Similarly, Concord-2 was another cross-border data project that did a comparative study of factors affecting cancer survival. The Concord-2 project analyzed data from over 270 cancer registries in 61 countries to identify the global differences in cancer survival rates and the main reasons for the differences.[126]

Political/regulatory impetus and impediments in India

The Indian government is also attempting to build a unified database for health. The National Health Stack (NHS, similar to India Stack) is an ambitious plan to create an electronic national health registry that would function as a master database covering 1.3 billion Indians. Another component of the NHS is the Federated Personal Health Records (PHR) framework that would make health data available for medical research. Simultaneously, immense efforts are being made to formalize and digitize Indian healthcare data.[127, 128]

The political and regulatory framework in India is supportive of developing data infrastructure and deploying AI. There are several encouraging signs from the Indian government. In the national AI strategy document prepared by the NITI Aayog, healthcare is a priority sector and the authors advocate the use of robotics and the Internet of Medical Things. However, there isn't much clarity on how this will be implemented.

Apart from the strategy document, the Indian Ministry of Health and Family Welfare (MoHFW) is also running several e-governance initiatives for the digitalization of the healthcare sector in India, which includes creating a National eHealth Authority (NeHA), Electronic Health Record Standards and an Integrated Health Information Program.

Yet lack of clarity on the processes on the ground still hinders the adoption of AI. There is no regulatory authority for AI in healthcare. For start-ups that wish to bring their innovations to market, there is a lack of appropriate certification mechanisms. Further, they face difficulty in conducting clinical trials as there are no clear regulations to adhere to.[129]

Potential for cross-border exchange of health data

A look at the business landscape in India tells us that all actors are increasingly trying to incorporate AI in healthcare. The growing network of start-ups and the interest shown by MNCs in working with Indian hospitals indicate that the private sector and foreign players also recognize the potential of India's data. The Indian government, in its intent and through its actions, has also attempted to pilot projects in AI. The government's investment in research and its creating robust data infrastructure (like the National Health Stack) indicate that the country is geared for a transformation of its health sector. Further, bilateral agreements between Germany and India identify the development of AI for healthcare as a priority sector.[130]

Our interviews confirmed, however, that the successful models are currently the ones through which companies develop and deploy AI products in collaboration with established hospitals in India. These projects – if not embedded in global MNC efforts – are necessarily very limited in scope and are mainly driven by individuals on the ground.

The major challenge is the absence of EHRs in a fragmented health system. Where EHRs exist, maintenance and interoperability remain a challenge. There is a need for funding and multi-stakeholder cooperation to build capacity in hospitals and to boost interoperability. If India achieves its targets and implements EHR standards, there is great potential to grow the health data repository.[131]

5.2. Deep dive II: Data for AI and India's retail sector

The retail industry is rife with data and has been a leader in the use of data to increase business efficiency and consumer satisfaction. As an early adopter of technology[132] and with information about every stage of its business operations and consumer activity collected and connected, "retail e-commerce" has emerged as a global phenomenon, with over $3.5 trillion in B2C sales in the year 2019.[133] There are several industries that operate within it, including logistics, supply chain and vendor management, financial transactions, and advertising, among others.

The internet and mobile-app-based operations and the digital linkages with several industries have not only streamlined the collection of data, but have also provided access to several secondary data sources. This extensive data ecosystem is supported by a mature consumer data and analytics industry that is increasingly utilizing this data for a wide variety of AI applications. The benefits of AI for retail has been recognized, something that is reflected in the investment made by the sector to develop it. A study by Juniper Research estimated that $3.6 billion was spent on AI by global retail

in 2019, and the amount is expected to increase to $7.3 billion in 2022[134] and $12 billion in 2023.[135] A UNCTAD study[136] found that, while most online shopping is done by domestic suppliers, there has been an increase in cross-border purchases. The internationalization of e-commerce is reflected in the numbers, with the share of cross-border online shoppers increasing from 15% of total online shoppers in 2015 to 21% in 2017.

Cross-border data exchange is a necessary component of global e-commerce services: Executing an international order would be impossible without sharing data on customer orders with third-party vendors who provide goods or logistic services like transport and delivery. Additionally, cross-border data flows could serve as a means to employ AI in retail. Operations and customer data could serve as the input required to train machines and automate various tasks, such as smart warehousing and chatbots for 24/7 customer servicing. For firms with limited access to the data required to train AI, this cross-border exchange could result in significant gains for developing smarter AI for their local context.

The retail-tech landscape in India

The Indian retail and e-commerce sector is ballooning and, with it, the amount of data collected. With a large population and increasing penetration of smartphone and internet access, India was the ninth largest e-commerce marketplace in 2019 and reported the second largest growth worldwide, with a year-on-year increase of 31.9% between 2018 and 2019.[137] The size of the market was $46 billion in 2019, and it is expected to grow to $200 billion by 2026.[138] E-commerce start-ups received $4.7 billion in funding in 2017, which was 33.9% of all start-up funding that year.[139] Moreover, in 2016, $9.1 billion from India was spent on cross-border shopping.[140] The immense data from India generated as a by-product of routine activities could be a valuable resource for training AI models that can automate various consumer services and operational processes.

Data actors

There are several stakeholders in the retail value chain that produce data for AI or use this data to create AI applications. The two main actors are retail firms (ranging from small firms to Flipkart/Amazon), vendors in the supply chain (sellers, logistics providers, warehousing firms) and data sellers (social media platforms, mobile and internet service providers, data vendors, etc.)

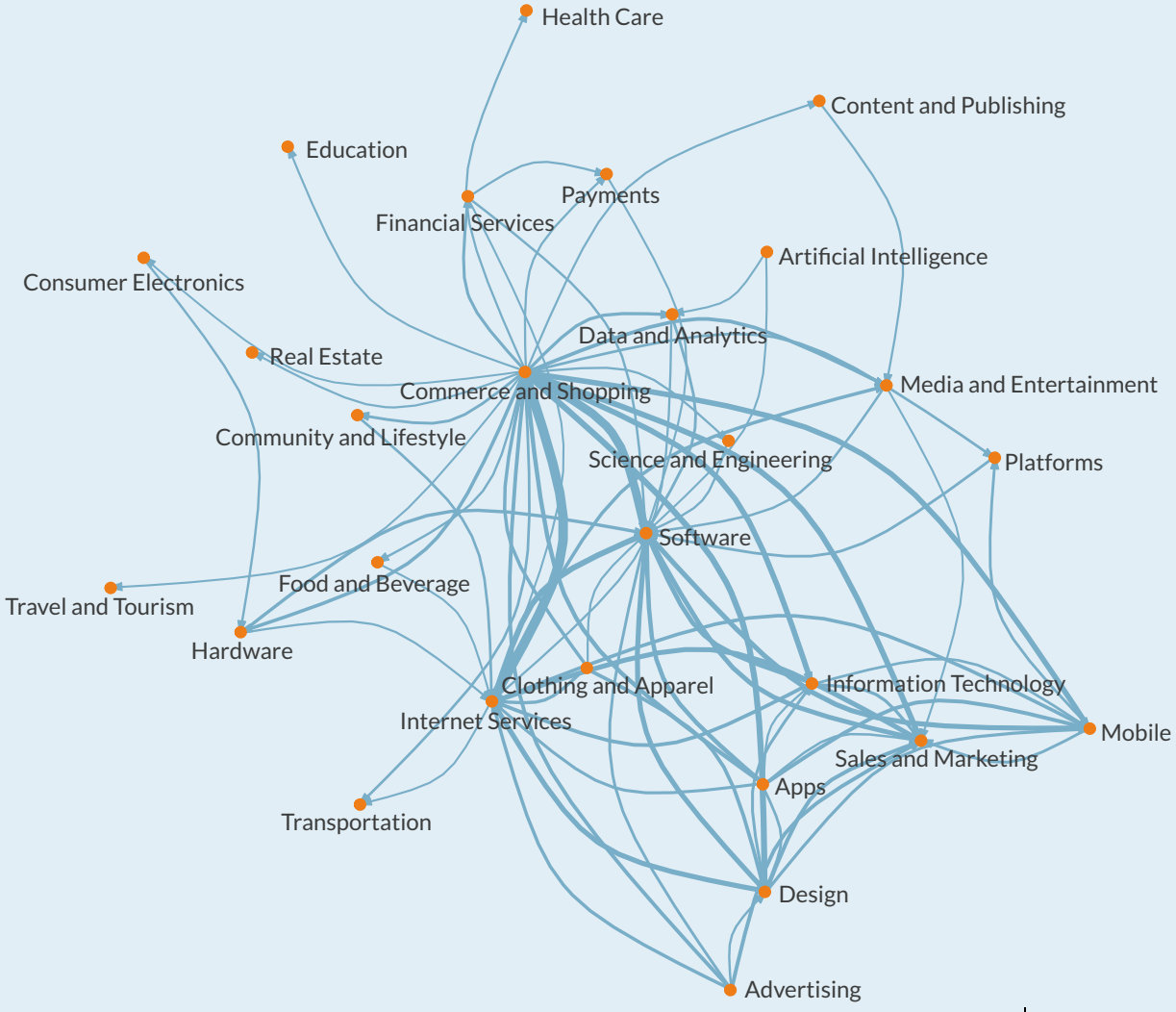
Two e-commerce giants in India occupy 80% of the market. Flipkart is the leading player, capturing 47% of the market with over 100 million users and generating 10–15 terabytes of data on an average day.[141] This is closely followed by Amazon India, which has a market share of 33%.[142] Smaller players include Snapdeal, PayTM and EBay.

According to our analysis, 160 core-AI firms were active in retail in India in 2019, i.e. they employ AI at the heart of their product offering. Several major sectoral players like Snapdeal (e-retail platform) and Myntra (fashion retail) also employ AI for chatbots and product recommendations; however, it is not a part of their core business. The network chart illustrates the various sectors that retail technology companies engage with (see figure 7). The sectors that engage most with retail-tech (consumer goods, clothing and apparel, consumer electronics) are internet services, sales and marketing, mobile and IT.

Data projects and sources

Despite all the cross-border data exchanges that are happening with the uptake of e-commerce, we were not able to identify any major cross-border data project in this sector. In fact, our interviews showed that even within internationally active Indian retailers, there is limited or no exchange of data for AI between the different entities in countries. While there is intra-firm investment by giant e-commerce firms in research, the extent to which data is used remains at a piloting level.

FIGURE 7 **SECTORS AND TECHNOLOGY THAT FIRMS IN RETAIL ENGAGE WITH**
Retail firms include commerce and shopping, clothing and apparel and consumer electronics. The thickness of the line is indicative of the number of engagements between retail firms and other sectors and technologies. The top sectors and technologies that retail engages with include software, IT, sales and advertising, design and mobile.



Source: Analysis by CPC Analytics.

BertelsmannStiftung

There are few data sources or projects in the public domain. Apart from semantic databases for training chatbots, European e-commerce company Zalando (based in Germany with operations in 17 countries) has developed a public database called Fashion MINST to train algorithms for image recognition of shopping goods.[143]

One key source of data for retail is telecom and internet service providers. Further, there is also a "shadow economy" of personal data stores from "data brokers"[144] that contain huge amounts of personal data. However, the legality of this is questionable, and our interviews indicate that the market for purchased "third-party data" for AI is extremely limited.[145] This confirms research on the data broker market in India.[146]

Political/regulatory impetus and impediments in India

The political and regulatory framework in India for retail and e-commerce, including cross-border retail, is supportive on paper, but uncertain in implementation. The Indian government allows 100% FDI in the e-commerce marketplace to attract the participation of foreign players. The proposed PDP bill in India is expected to encourage cross-border transfer of data with an "adequacy test" and "comparable level of protection test" for personal data. This is expected to enable smooth two-way flow of data, especially with countries that also have a strong data protection system (such as countries in Europe). However, in the absence of a proactive data protection authority (DPA) in India, the burden would be on the retailer to ensure that data protection is sufficient.[147]

India is also in the process of drafting an e-commerce policy[148] in which data is deemed a "national asset." It is hinted that there might be stricter checks and balances on cross-border flows of data that e-commerce firms operating in India might have to adhere to. However, how this will be operationalized has not been addressed.

As retail operates in a competitive market, the Competition Commission of India (CCI) monitors the activities of retail firms to ensure that there is no appreciable adverse effect on competition in India. A report by the CCI on e-commerce[149] noted that some competition issues could arise, such as giving preferential treatment to the platform's own products, artificially suppressing prices, influencing the customer's choice by modifying search rankings, and a lack of credibility of reviews. It further observes that because platforms have a superior bargaining position, this could lead to "unfair" and inconsistent contract terms, where the sellers have little power over pricing and discounts. Based on several of these issues arising in a highly unbiased marketplace, the CCI recently ordered an investigation into alleged competition law violations by Amazon and Walmart's Flipkart. The accusation is that exclusive arrangements between mobile phone brands and e-commerce platforms "merit an investigation," as do allegations of e-commerce companies giving preferential treatment to certain sellers.[150] The investigations are underway, indicating that the CCI exerts significant influence when monopolies and malpractices are present, a positive signal for new players wishing to enter the market.

Potential for cross-border exchange of retail data

In theory, e-commerce offers tremendous potential to deliver data relevant to AI development in other countries as well. Customer choices for certain products might differ greatly across countries, but there would be significant overlaps for several products.

However, the data from this market that is useful is kept mostly "within the firewalls" of companies and used to increase operational efficiency and predict consumer behavior. This data is a strategic asset and constitutes the base for competitive advantages. We therefore conclude that the potential of cross-border exchange between India and Germany to support AI development in the German retail sector is low.

6. The baselines of German-Indian data/AI collaboration

While data is a vital input for AI, many other factors influence the environment that facilitates innovation in AI in general and deployment of machine-learning models more specifically. This includes the availability of a talented and AI-trained workforce, a well-developed digital infrastructure, a vibrant research community in related fields, such as computer science and statistical modelling, and a dynamic innovation landscape – ranging from innovative start-ups to supportive incubators and ultimately to investors.

6.1. Locating Germany and India in the global AI landscape

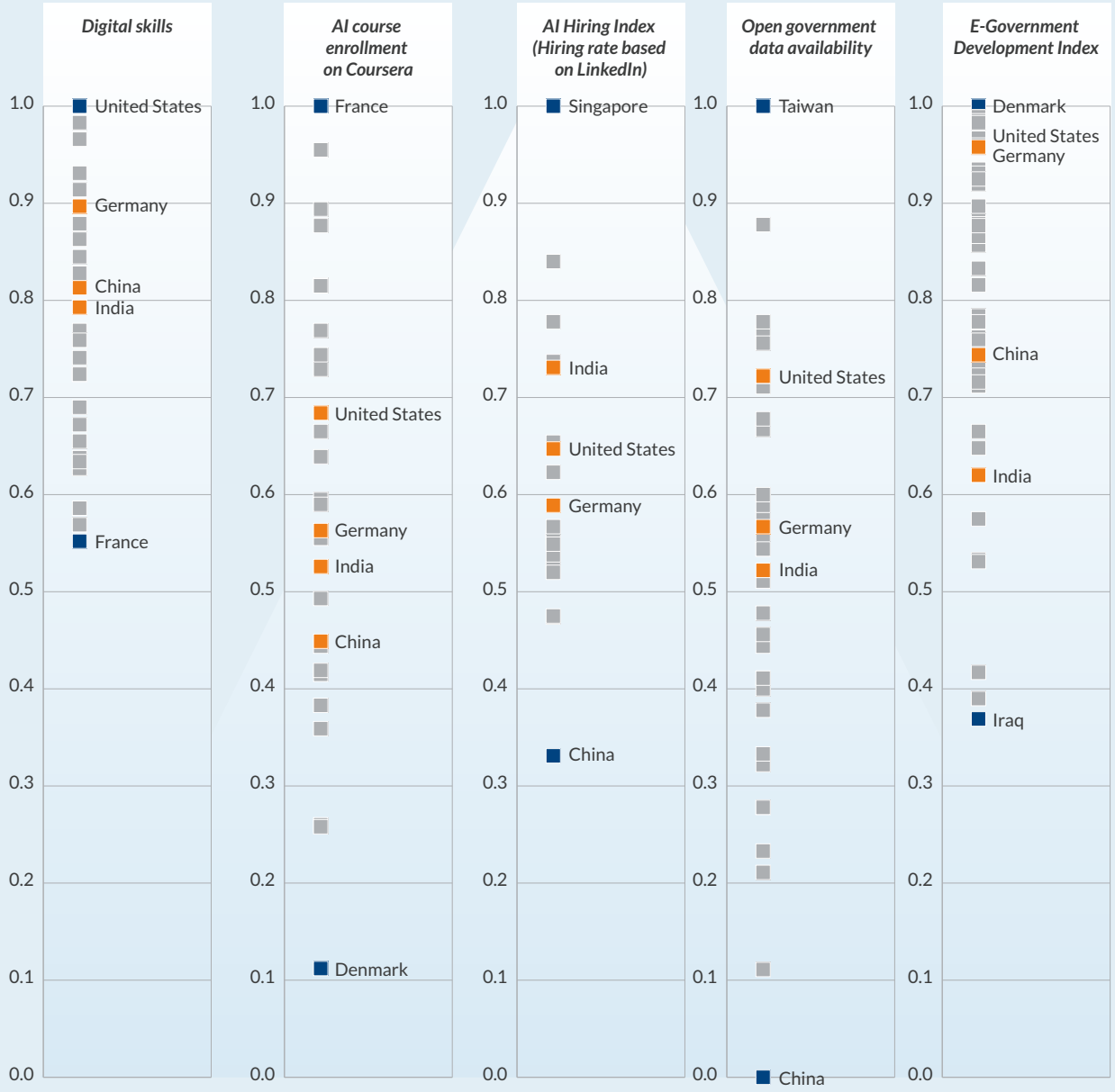
AI has been high on the policy agenda in both Germany and India. The two countries are seen as second movers when it comes to establishing AI as a strategic policy goal.[151] There is now a variety of composite indicators that attempt to quantify the multidimensional nature of the "AI readiness" of governments[152] or countries[153] as well as "global AI vibrancy."[154] For the purpose of this study, a selection of individual indicators appears more helpful to achieve an overview of where Germany and India stand relative to other large economies (top 50 by GDP). While such country-level comparisons necessarily fall short of capturing the complexity of AI innovation, they nevertheless help to see where closer collaboration between Germany and India could support progress in both countries.

The chart on the next page summarizes 13 indicators, allowing a glimpse of the status quo of multiple factors relevant to AI. For each indicator, we show the performance of countries relative to the leading country in that dimension. Our sample covers the 30 largest economies by GDP, but not all countries are available for all indicators.

Talent, skills, jobs

India ranks second after Kenya among the lower middle-income countries when it comes to digital skills, but is far behind the US and other European countries including Germany. When compared with these countries for "Proportion of AI-related course enrolment to all courses," India surges ahead of some of the European countries, like the Netherlands and Denmark, as well as China and Singapore in Asia. The higher percentage of enrolment on Coursera in India shows the anticipated demand of these skills in the market as well as the readiness of talent to adapt to the change. The AI Hiring Index is an indicator of the dynamism in the demand for AI talent in different countries as compared to the baseline situation in 2015-16. It looks at the proportion of candidates with AI-related skills on LinkedIn in a given month and measures how many of those had a new employer in the following month. India ranks fairly high, with an AI hiring rate in 2019 that is almost 2.5 times the one in 2015-16 – ahead of the US, Germany and France.

FIGURE 8.1 **AI-RELATED INDICATORS ON SKILLING AND PUBLIC SECTOR**
ALL INDICATORS RELATIVE TO THE LEADING COUNTRY (1 = HIGHEST, 0 = LOWEST).



Note: 50 largest economies by nominal GDP (2019 estimate by IMF). For several indicators, missing data for several countries limits the number of displayed countries. Source: Various, see appendix for details (section 9.2). Analysis by CPC Analytics.

BertelsmannStiftung

Research and development

We have used two major categories of indicators to measure the comparative standings of different countries in research and development: journals and patents. While the volume of publications shows the total output by country, the citations help us understand how influential the research is. In terms of the absolute output of journals, India ranks among the top three countries, even though the difference between the second country, the US, and India is substantial. When we consider

FIGURE 8.2 **AI-RELATED INDICATORS ON RESEARCH AND INNOVATION**
ALL INDICATORS RELATIVE TO THE LEADING COUNTRY (1 = HIGHEST, 0 = LOWEST).

Note: 50 largest economies by nominal GDP (2019 estimate by IMF). For several indicators, missing data for several countries limits the number of displayed countries. Source: Various, see appendix for details (section 9.2). Analysis by CPC Analytics.

BertelsmannStiftung

citations of these journals, India's ranking drops to number five, with the UK and Australia having more citations despite having fewer journals than India. When it comes to where most AI patents are published, the US leads by a wide margin: The number of AI-related patents in Germany is only a tenth of the number of published patents in the US – roughly on par with the situation in China. India is even further behind. When it comes AI patent citation, Germany ranks third – behind the US and Japan. India lags even further in this indicator than in the number of published AI patents.

FIGURE 8.3 **AI-RELATED INDICATORS ON PRIVATE SECTOR ACTIVITIES**
ALL INDICATORS RELATIVE TO THE LEADING COUNTRY (1 = HIGHEST, 0 = LOWEST).

Note: 50 largest economies by nominal GDP (2019 estimate by IMF). For several indicators, missing data for several countries limits the number of displayed countries. Source: Various, see appendix for details (section 9.2). Analysis by CPC Analytics.

BertelsmannStiftung

Start-up and investment in AI

India and Germany are among the top five countries in terms of the number of registered AI-related start-ups. It is remarkable to see that India and Germany are ahead of France and China (however slightly) when it comes to the number of AI start-ups. A different finding emerges when one looks at the private investments related to AI. The US and China lead on this indicator by a wide margin. Germany and India are at similar levels, but quite substantially behind the UK, Israel and Canada.

Automation

This dimension is particularly interesting from a data perspective: Machine-generated data has great potential to improve supply chain, manufacturing and distribution processes. Globally, Germany is third, with 366 installed robots per 10,000 manufacturing workers. Singapore and South Korea are the global leaders as measured by the installed robot base. India, with just three installed robots per 10,000 manufacturing workers, is still far behind. Again, the comparison to China is telling, given its robot density of 140 robots per 10,000 manufacturing workers.

6.2. Estimating the current volume of the German-Indian AI-related economy

Unsurprisingly, sizing the economic potential of a strategic German-Indian collaboration on data for AI is a challenge. Given the absence of any figures on this potential, we provide here two rough estimates based on trade in ICT-enabled services and on the size of the domestic digital economies (see table below).

The trade volume between Germany and India in potentially ICT-enabled services had reached €4.6 billion by 2018.[155] As most data exchanges involve some digital services (e.g. cloud, analytics, platforms), this volume can be seen as a reasonable starting point. Given that the current AI-related "investment intensity" of firms is around 10–20% of their overall digital investment budget,[156] it is reasonable to assume a baseline for the economic potential of a German-Indian data exchange for AI of around €460 million to €920 million annually.

An alternative approach is to look at the size of the domestic digital economies in both countries. The combined digital economies of Germany and India are estimated to be around $327 billion annually.[157] Most applications of AI will involve digital firms – even if only as support for larger companies (e.g. large car OEMs). Given the countries' existing bilateral exports as a share of total exports, the economic potential of the bilateral trade between the digital economies would be around $8.27 billion. Applying an "investment intensity" of 10–20% once again, the economic potential of the data exchange for AI between the two countries could easily reach €763 million to €1.522 billion.[j]

TABLE 1 **ROUGH ESTIMATES OF THE GERMAN-INDIAN AI-RELATED ECONOMY**

	Estimated volume in million euros annually
Based on ICT-enabled services	460 – 920
Based on digital economies	763 – 1,522

Note: Estimate based on digital economies converted to euros using the average exchange rate for 2017 ($1 = €0.923).

BertelsmannStiftung

6.3. Status quo and pathways towards German-Indian data collaboration for AI

Status quo

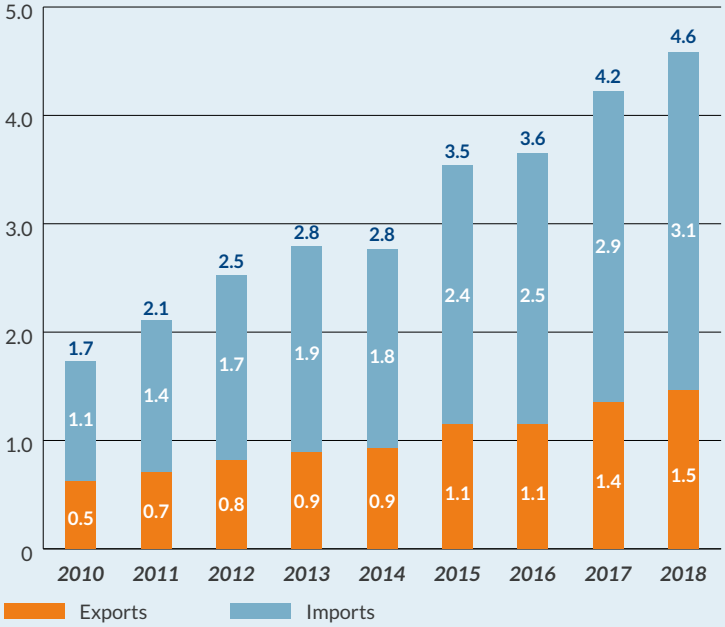
In November 2019, collaboration on AI and digitalization was a major topic at the fifth Indo-German Inter-Governmental Consultations. Both heads of government expressed their strong willingness to work on a "Digital Partnership to intensify regular interaction and coordination towards collaboration on the next generation technologies."[158] Such a collaborative partnership should aim at "leveraging advantages on each side recognising increasing integration of hardware and software in developing IoT and AI solutions for societal benefits."[159] The sectors health, mobility, environment and agriculture were mentioned in particular as good starting points to build on. Among other things, the two leaders welcomed the formation of a digital expert group made up of German and Indian companies and research institutions.[160]

j The USD values were converted into euros using the 2017 average exchange rate published by the US Internal Revenue Service (IRS). https://www.irs.gov/individuals/international-taxpayers/yearly-average-currency-exchange-rates

FIGURE 9 **TWO OPTIONS TO ESTIMATE THE BASELINE OF A GERMANY-INDIA DATA COLLABORATION**

Sizing the potential of German-Indian data exchange

Trade volume in potentially ICT-enabled services between India and Germany
In billion USD

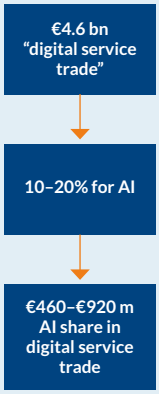


Source: stat.oecd.org, retrieved March 19, 2020. Categorization of "potentially ICT-enabled services" from Grimm (2016). Analysis by CPC Analytics.

Why is trade in potentially ICT-enabled services a good starting point?

- Most data exchanges involve services (e.g. software, analytics, etc.)
- Actual market for data trade (incl. ownership transfer) is very small.

Calculation



2018 trade volume between India and Germany on potentially ICT-enabled services is €4.6 billion.

"Various corporate surveys conducted by MGI and McKinsey suggest that AI intensity within digital (AI-related investment out of total digital investment) is around 10 percent today. If this ratio grows in line with the overall pace of AI adoption, it could reach around 35 percent by 2030."

Source: ITU (2018).

Sizing based on GVA of the digital economies

Currently, India's digital economy generates about $200 billion of economic value annually — 8 percent of India's GVA in 2017–18 — largely from existing digital ecosystem comprising of IT and business process management (IT-BPM), digital communication services, e-commerce, domestic electronics manufacturing, digital payments, and direct subsidy transfers.

Source: MEITY (2019).

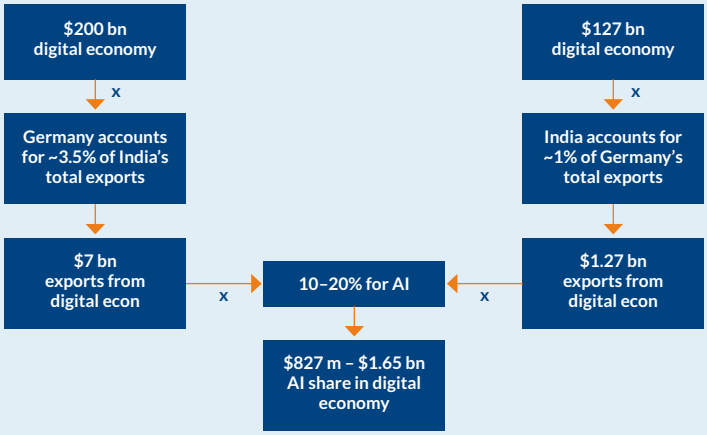
The German ICT sector was able to increase its gross value added in 2017 to €108 bn – an increase of 4% compared to the previous year. Since 2010, gross value added – the value of the goods and services produced less intermediate consumption – has increased by a total of €30 bn.

Source: BMWi & ZEW (2018).

Why should the "digital economies" in both countries be starting points for a sizing of the opportunity?

- Most applications of AI will involve digital firms.
- Even if the firms applying AI come from all sectors, it should be expected that services from digital firms are needed and these revenues will show up in the "digital economy".

Calculation



That both countries have every reason to speak about their digital economies becomes obvious in light of their trade in digitally deliverable services.[161] These flows represent about 20% and 16% of total service imports in Germany and India, respectively. Even more strikingly, both countries export more than 30% of their services in a digital way (for India it is even 34.2% vs. the OECD average of 28%).[162]

German multinational corporations have invested for decades into production and development facilities in India. The annual FDI flows from Germany to India were around $900 million between 2014 and 2019.[163] The Electronic City in Bangalore is one such place where Bosch or Siemens have also invested in data analysis capacities. The Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI) for example – founded in August 2017 – carries out basic and applied research in the field. Conversely, Indian firms have increasingly started to invest in German firms – even though the relative size of the investments is still limited. The UK remains the premier preference for Indian investors.[164] It is interesting to note that investments in software firms represented the largest share of Indian FDI projects in Germany between 2010–2017 (23%).

There is also potential from a research perspective. In November 2019, the Indian Ministry of Science & Technology and the German Ministry for Education and Research signed a Joint Declaration of Intent for Joint Cooperation in R&D on AI. Both countries agreed to hold a workshop in Berlin in 2020 "to identify areas of mutual interest," organized by the Indo-German Science and Technology Center (IGSTC).[165]

This ties into previous technology cooperation between the two countries: The IGSTC was opened in 2010 in Gurgaon and represents a unique project in Germany's international cooperation efforts. The IGSTC supports R&D projects in both countries and promotes networking between German and Indian researchers through workshops and symposia.[166] The projects mainly follow the "2+2 format," i.e. collaborations that involve an industry partner and a research institute on both sides.[167] However, AI has only started to play a more significant role since 2018. There are some projects linked to data and machine learning (e.g. the Translearn project focusing on AI and robotics between Tata Consulting Services and IIT Kanpur, on the Indian side, and Kuka and Karlsruhe Institute of Technology, on the German side).[168]

The Indo-German Centre for Sustainability (IGCS) at IIT Madras is another research partnership formed by the two countries and is focused on sustainable development in Germany, India and South Asia.[169] It should also be mentioned that Indian students are the second largest group of foreign students in Germany, behind China but ahead of Russia. Currently, 20,800 Indians are studying in Germany.[170]

As the paragraphs above indicate, there are certainly starting points for an Indo-German data collaboration for AI. Also, there was general agreement among the experts interviewed for this study that data sharing for AI development will be necessary in the future. However, when asked whether there are any examples they could give showing where such data exchange or data sharing had been useful for their past work on AI and analytics, the interviewed experts could not name any that went beyond their own company. Narrowing our research to sectors that appeared most interesting for data exchanges with Germany (and not predominantly useful in the Indian context), we were able to generate a differentiated view of the following four sectors:

Future pathways

But what might concrete projects and initiatives look like that would allow Indian and German actors to realize closer ties on AI development? Based on our research and the expert interviews we see two action areas: first, initiatives that support the construction and maintenance of German-Indian data sets suitable for AI; second, projects to increase expert-level collaboration for AI development.

TABLE 2 **HOW DID THE INTERVIEWED EXPERTS ASSESS THE STATUS AND POTENTIAL OF CROSS-BORDER DATA SHARING FOR AI?**

	Is data for AI available?	Is the data shared with other actors?	Is the Indian data transferable to German context?
Health	Hospitals are the predominant source for AI-relevant data, which is still highly fragmented.	Health data is generally not shared with entities outside of India. However, research collaborations exist (not specifically for AI).	Image data from diagnostic devices is transferrable to the German context. The usefulness of data from other devices like wearables, etc. is unclear. Health-system data is unlikely to be useful in Germany.
E-Commerce	Consumer behavior data is available.	Data remains within corporations and companies. Even within the same company it is hardly shared.	Theoretically, much of online behavior data could be transferred to a German context. True value is unclear.
Manufacturing	Data analytics in production processes is more and more widespread, but the usefulness of this data for AI is assessed to be limited at this stage.	Data is barely shared even within companies.	Theoretically, machine data is transferrable, but production processes differ quite substantially from Germany.
Open Government	The currently available government data is not useful for AI development.	More and more data is shared publicly.	Only small subsets of data would be comparable (e.g. smart cities).

Source: CPC Analytics. BertelsmannStiftung

Until there is clarity on the future of India's data protection law, meaningful progress on a larger scale seems very unlikely. There is an open window of opportunity as the Indian data protection regulation has not been finalized and European efforts might fall on fertile ground. If they were to succeed, a reassessment of the situation would be required, one that also draws on learnings from bilateral lighthouse projects. As AI and international data governance are both still nascent areas, it is difficult to assess whether further policy action will be required two years from now or whether non-governmental actors and initiatives will suffice at that point to facilitate generation of useful data for AI cooperation.

With regards to **increasing the availability of data for AI development**, the global discourse has (unsurprisingly) not settled on a preferred approach. In the absence of universally binding and enforceable national and international rules on data sharing, the contemporary discourse has looked into alternative governance mechanisms for data sharing and use. The Open Data Institute suggested the term "data institutions" as an umbrella term for "organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims."[171] One promising and widely discussed "vehicle" is the data governance model of "data trusts,"[172] which can be described as legal entities that manage data or the rights to data.[173] Data trusts represent an opportunity to build an agreement on key aspects of a specific collaboration: Who is the beneficiary from the products emerging from data sharing? Who has access to the data? Who protects the rights of those sharing the data?

In the health sector, for example, several initiatives in the UK have started to explore institutional setups to share specific data for research purposes. For example, the INSIGHT Health Data Research Hub will bring together anonymized data from eye scans and images and from advanced analytics and make it available to the National Health Service (NHS) and academic and industry researchers. "It aims to unlock new insights in disease detection, diagnosis, treatment and personalised healthcare."[174] A key in the setup is to engage with all stakeholders (including

people whose data is shared) and accommodate their needs and anxieties around data sharing. However, given the novelty of most arrangements, evidence is scarce of successful data trusts on the national level. Internationally, there is hardly any standardized effort to share data widely – with genomic data being an exception.

However, answering the questions involved in the generation and maintenance of such a data trust goes beyond the power of most companies (whether start-up or multinationals). Thus, policy makers wanting to promote data collaboration have a role to play in initiating and guiding such a process. Given the complexity of such arrangements, policy makers addressing the subject of data collaboration for AI between India and Germany need to consider starting with data sources that are less linked to personal data. For example, while X-rays of lungs are personal data, they might be more easily anonymized than complete healthcare records.

The possibility of **increasing expert-level collaboration for AI development** exists even before data sharing capacities are implemented. Two ways could be explored: On the one hand, several experts interviewed expressed hesitation to share data beyond borders following enactment of the GDPR – the uncertainty over potential investments in data protection seems too daunting. Such uncertainty could be addressed by deliberately supporting an expert community working collaboratively on AI projects. In publicly visible lighthouse projects, both governments could issue calls for evidence on how to utilize cross-country datasets for common AI development, while at the same time identifying data protection experts who can ensure compliance with the GDPR. The power such initiatives can have for mobilizing a wider audience for a technical challenge with public relevance was demonstrated when the German government held the hackathon #WirVsVirus, which included more than 1,500 projects. These lighthouse projects would contribute in two ways: First, AI experts would be incentivized to invest in knowledge around data protection in the European Union – making it possible to again view Germany/Europe as an innovative partner. Second, it would allow a bottom-up identification of problems in India and Germany that are addressable with data and AI.

On the other hand, Germany and India could support bilateral platforms in utilizing "distributed learning" or "federated learning" approaches between companies and researchers. The idea is that AI algorithms can be trained on local data, selected parameters of the trained models can be shared across borders, and the precision of the individual models can be increased. In these approaches, neither input data nor results need to be shared. The German GAIA-X initiative highlights the use of such approaches in the healthcare sector to allow AI-relevant data to be processed locally without creating a unified database – which would offer the potential for misuse.[175] Creating a common platform and a working approach that take advantage of the technical opportunities offered by federated learning between India and Germany would require a significant up-front exchange of experts. Ideally, research/corporate actors would be financially incentivized to experiment on selected topics.

7. Synthesis

Our analysis started with the observation that, to compete successfully in AI, the data space in Germany/the EU needs to expand access to high-quality data sets, in keeping with an open, yet heavily regulated approach to cross-border data transfers. Data for AI is (still) a precious good and our analysis has aimed to assess the potential of a balanced data cooperation among equals, namely India and Germany/the EU.

Indeed: Our analyses of India's digitalization efforts and of the private sector landscape point towards a rapidly evolving environment for digital innovation. India's government has launched a series of initiatives to support not only the digital economy, but also the country's digital development. Useful data has increasingly been collected and the country certainly has the talent pool and industry to make use of this data.

One major barrier for large-scale data exchange is regulatory uncertainty: Our analysis of the German/European requirements has shown that the GDPR sets standards for cross-border

FIGURE 10 **POTENTIAL OF INDIA-GERMANY DATA EXCHANGE FOR AI**

Three dimensions of potential

	Data potential	Ecosystem	Regulation
Needs	• High-quality, large data sets • Companies and organizations able and willing to share data • Shared trust and values	• Highly trained data experts • Specialized research landscape • Vibrant innovation ecosystem • Political stewardship	• Mutual adequacy of data protection regimes • Guidelines for data exchange for AI experimentation
India: Key findings	✓ India Stack helps digitalizing economic transactions (a health stack is planned) ✓ Substantial efforts in utilizing health data ✓ Vibrant e-commerce and digital payments landscape ! Limited manufacturing data ! Health data is often fragmented or not available	✓ India's IT-sector and universities provide a promising talent pool ✓ 890 "core-AI" firms based in India (90% founded post 2011) ✓ Political momentum towards AI and computing infrastructure ! Low recognition of research and development from India ! Ecosystem mostly limited to urban centers (tier 1 and tier 2)	✓ Indian AI Strategy in place ✓ Draft Personal Data Protection Bill to be agreed on in 2020 ✓ "Innovation sandboxes" for AI development ! Uncertainty over definitions in the privacy bill as well as government intervention ! Uncertainty over de facto enforcement of the PDP Bill

Source: CPC Analytics.

BertelsmannStiftung

transfers of data. For systematic data sharing across borders, India would ideally receive the status of an "adequate jurisdiction" from the European Commission. At present, however, our analysis of the Indian regulatory situation shows that final clarity on most data protection questions will only be achieved by 2022 – once the data protection authority (DPA) is established.

Other major barriers became apparent in the AI landscaping and the deep dives. Data in the health sector is still fragmented and obtaining a unified data set entails tremendous costs. Again, there are a number of promising plans to change this situation, but nationwide implementation will take time. In the e-commerce sector, data is unlikely to be shared given that it represents a strategic asset for the firms involved. Manufacturers are experimenting with AI, but in most production contexts AI development has not yet been prioritized over other optimization approaches.

These major barriers should not lead to the conclusion that efforts to build and foster meaningful cooperation in AI should be postponed. Rather, they need to be expanded (see Recommendations). Our analysis highlights the potential of cross-border data collaboration between India and Germany/the EU. In the immediate future (at least until 2022), however, regulatory uncertainty will persist and key sectors will still grapple with obtaining reliable data.

8. Recommendations

The goal of this study was to explore the potential for data exchange between India and Germany for developing AI. The potential for a systematic "data-for-AI"-partnership in the immediate future is limited. **Against this background, a major political effort to foster cooperation in data for AI might not be the most promising candidate for the bilateral agenda for now**, as it would bind a lot of policy makers' time and resources for a rather long period and for a mission that is theoretically appealing but not yet fully proven on a practical level. This obviously is not meant to discourage efforts and initiatives on a private level, which to some extent might not even need policy makers' support. But until there is clarity on the future of India's data protection law, meaningful progress on a larger scale, beyond lighthouse initiatives or mechanisms to circumvent critical data transfer issues (like federated/distributed learning), seems very unlikely. **Championing an Indian PDP that is likely to gain adequacy status in the EU thus seems the single most important high-level topic that policy makers should focus on for now.** There is an open window of opportunity as the Indian data protection regulation is not finalized yet and European initiatives might fall on fertile ground.

The following approaches have the potential to improve relations and facilitate exchange, even before an adequacy status is decided upon, but are of considerably smaller scope:

Establish consensus on strategic goals for AI leadership:

- **Ethics in data sharing and AI:** Germany and the EU should build on their expertise on "ethical AI" and how to put guidelines into practice (compare, for example, the VCIO model by VDE and Bertelsmann Stiftung, or the membership in the Global Partnership for AI). Building on the work of the Indian Task Force on Artificial Intelligence, the dialogue on these rules would constitute a platform for building trust in practices around data collection and data exchange in India and Germany. Civil society organizations in India and across the EU play a vital role in the debate around AI and data protection. These actors should not only be considered in these dialogues, but should actively be enabled to contribute.

Establish trust in the EU/Germany as a visible AI leader in India:

- **Increase the availability of data for AI development:** Germany, the EU and India should build publicly funded common reservoirs for data on challenges of public interest (e.g. mobility, health) to lower the market entry barrier for AI start-ups and other companies in these areas. A bilateral task force could identify sources of these data on both sides and ensure anonymization of the data before offering the data set to interested parties. Different data governance models should be explored to ensure data stewardship and accommodation of different interests.

- **Increase expert-level collaborations for AI development:**

  - Germany and the EU should foster publicly visible projects that allow companies on both sides to see that cross-border AI projects can be implemented "despite" and in full compliance with the GDPR. For that purpose, lighthouse projects in sectors of strategic interest in both regions (healthcare/mobility) could be launched. These projects can be supported by European/Indian privacy and data experts to ensure compliance with Indian and European data protection regulation.

  - Germany and the EU should approach India to build a federated learning infrastructure.[k] Given the uncertainty over cross-border data transfers between Germany/the EU and India, technical solutions can be explored that allow training of AI models on Indian data without the data leaving the country.

k There are several starting points in Germany: First, the research platform "Learning Systems" – initiated by the Federal Ministry of Education and Research (BMBF) in 2017 – has worked on multiple aspects of an AI-infrastructure in Germany (https://www.acatech.de/projekt/lernende-systeme-die-plattform-fuer-kuenstliche-intelligenz/). Second, the Berlin-based AI firm XAIN AG has worked on developing a platform for federated learning (https://www.xain.io/assets/XAIN-Whitepaper.pdf). Third, a working group at the Max-Planck-Institute for Intelligent Systems in Tübingen, for example, explores ways to share locally learned models with others without leaking sensitive information about private data (see https://privacy-preserving-machine-learning.github.io/index.html). In India, the Indo-German Research Centre for Intelligent Transport Systems at IIT Karagpur would be a good starting point for mobility-related data. Wadhwani AI could be a starting point for the discussion of a common learning infrastructure.

9. Appendix

9.1. Definitions

Digitization/Digitalization/Digital Transformation: "Digitisation is the conversion of analogue data and processes into a machine-readable format. Digitalisation is the use of digital technologies and data as well as interconnection that results in new or changes to existing activities. Digital transformation refers to the economic and societal effects of digitisation and digitalisation."[176]

AI Systems: These are systems that "display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)."[177]

Machine learning: "This is a set of techniques to allow machines to learn in an automated manner through patterns and inferences rather than through explicit instructions from a human. ML approaches often teach machines to reach an outcome by showing them many examples of correct outcomes. However, they can also define a set of rules and let the machine learn by trial and error. (…) These range from linear and logistic regressions, decision trees and principle component analysis to deep neural networks."[178]

(Deep) neural networks: "The real technology behind the current wave of ML applications is a sophisticated statistical modelling technique called 'neural networks. This technique is accompanied by growing computational power and the availability of massive datasets ('big data'). Neural networks involve repeatedly interconnecting thousands or millions of simple transformations into a larger statistical machine that can learn sophisticated relationships between inputs and outputs. (…) Finally, deep learning is a phrase that refers to particularly large neural networks; there is no defined threshold as to when a neural net becomes 'deep'."[179]

Federated learning: "An approach to machine learning for training a central, shared model using data that is distributed across multiple locations rather than available centrally. It has applicability where all the training data is not available in the same place or at the same time or where it is not possible or desirable to bring the training data into a central location. It allows the data to be used where it exists without the need to remove it from its location (i.e. a mobile phone or other device) and then uplinks the learnings back into the central model without having to send the actual data back."[180]

9.2. Indicators used in section 6.1

Indicator	Source	Description
Digital skills	WEF Global Competitiveness Report 2018 Year: 2018	Digital skills among active population: Response to the survey question: "In your country, to what extent does the active population possess sufficient digital skills (e.g. computer skills, basic coding, digital reading)?" [1 = not all; 7 = to a great extent]
AI course enrollment on Coursera	Coursera Year: 2019	AI (% of total enrolment) — Coursera computes the fraction of a country's enrollments that are in courses teaching AI and related skills to measure the relative interest in AI content across the globe. Measuring this fraction over time allows us to see enrollment trends and where emphasis on AI is increasing or decreasing.
AI Hiring Index (Hiring rate based on LinkedIn)	LinkedIn: LinkedIn Economic Graph Year: 2019	LinkedIn Economic Graph — AI hiring rate is the percentage of LinkedIn members who had any AI skills on their profile and added a new employer to their profile in the same month the new job began, divided by the total number of LinkedIn members in the country. This rate is then indexed to the average month in 2015–2016; for example, an index of 1.05 indicates a hiring rate that is 5% higher than the average month in 2015–2016.
Open government data availability	Open Knowledge Foundation Year: 2016–17	The Open Data Index is an annual expert peer-reviewed snapshot of the country-level Open Data Census, and has been developed to help answer relevant questions by collecting and presenting information on the state of open data around the world – to ignite discussions between citizens and governments. The original index ranges from 0–100.
E-Government Development Index	UN e-governance survey by UN DESA Year: 2018	The EGDI is a weighted average of three normalized scores on the three most important dimensions of e-government, namely: (1) scope and quality of online services, (2) development status of telecommunication infrastructure and (3) inherent human capital.
Citations of journal articles on AI	Microsoft Academic Graph (MAG) Year: 2018	Total count of AI journal paper citations attributed to institutions in the given country
Number of journal articles on AI	Microsoft Academic Graph (MAG) Year: 2018	Total count of published AI journal papers attributed to institutions in the given country
AI-related patents	Microsoft Academic Graph (MAG) Year: 2018	Total count of published AI patents attributed to institutions in the given country
AI-related patent citations	Microsoft Academic Graph (MAG) Year: 2018	Total count of AI patents citations attributed to institutions in the given country
AI-related start-ups	Crunchbase Year: 2018	Number of AI start-ups per country as registered on Crunchbase
Total private investment in AI (2018–19)	CapIQ, Crunchbase, Quid Year: 2018	Total private investment in AI: Billions of US$
Autonomous military systems developed	Stockholm International Peace Research Institute (SIPRI) Year: 1959–2017	Number of autonomous military systems developed
Robot density	International Federation of Robotics (IFR) Year: 2016	Annual Installations of industrial robots (,000 of units)

10. References

1 OECD. 2019. Artificial Intelligence in Society Artificial Intelligence in Society. Paris: OECD Publishing. https://doi.org/10.1787/eedfee77-en

2 Varian, H. (2018). Artificial Intelligence, Economics, and Industrial Organization. NBER Working Paper Series, (July). https://www.nber.org/papers/w24839

3 Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. (2019). "Big Data and Firm Dynamics." AEA Papers and Proceedings 109: 38–42.

4 See for example Lee, K.-F. (2018) AI Superpowers: China, Silicon Valley, and the new world order. Houghton Mifflin Harcourt. And Chakravorti, B., Bhalla, A., & Chaturvedi, R. S. (2019) Which Countries Are Leading the Data Economy? Harvard Business Review (Online). https://hbr.org/2019/01/which-countries-are-leading-the-data-economy

5 Ciuriak, D. (2018). The Economics of Data: Implications for the Data-Driven Economy. Waterloo, ON Canada: Centre for International Governance Innovation (CIGI). https://doi.org/10.2139/ssrn.3118022

6 Bhaskar Chakravorti, Ajay Bhalla, R. S. C. (2019). Which Countries Are Leading the Data Economy? Harvard Business Review (Online), (January 24). https://hbr.org/2019/01/which-countries-are-leading-the-data-economy

7 Lee, K.-F. (2018). AI Superpowers – China, Silicon Valley, and the New World Order. Boston: Houghton Mifflin Harcourt.

8 Franke, Ulrike, and Paola Sartori. 2019. Machine Politics: Europe and the AI Revolution – Policy Brief. Brussels: European Council on Foreign Relations. https://www.ecfr.eu/publications/summary/machine_politics_europe_and_the_ai_revolution

9 European Commission. 2018. Artificial Intelligence for Europe COM(2018) 273 Final. Brussels: European Commission, p. 3. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625

10 For an overview of the EU's positioning see: Craglia, Max et al. 2018. Artificial Intelligence: A European Perspective – Joint Research Centre. Luxembourg: Publications Office of the European Union, p.17. DOI: 10.2760/11251

11 As can be read in the EU Data Strategy: "Fragmentation between Member States is a major risk for the vision of a common European data space and for the further development of a genuine single market for data. […] The value of data lies in its use and re-use. Currently there is not enough data available for innovative re-use, including for the development of artificial intelligence." European Commission. 2020. Communication COM(2020) 66 Final on "A European Strategy for Data." Brussels: European Commission, p. 6. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

12 The EU Data Strategy also puts an important focus on international cooperation beyond the Union's territory: "Building upon the strength of the Single Market's regulatory environment, the EU has a strong interest in leading and supporting international cooperation with regard to data, shaping global standards and creating an environment in which economic and technological development can thrive, in full compliance with EU law." Ibid., p. 23.

13 From a macroeconomic modelling perspective: Jones, C. I., & Tonetti, C. (2019). Nonrivalry and the Economics of Data. NBER Working Paper Series, (September). http://www.nber.org/papers/w26260. For an overview of alternative perspectives: Carriere-Swallow, Yan, and Vikram Haksar. 2019. 19/16 The Economics and Implications of Data – An Integrated Perspective. Washington D.C.: International Monetary Fund. https://www.imf.org/~/media/Files/Publications/DP/2019/English/TEIDEA.ashx

14 Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. Prediction Machines: The Simple Economics of Artificial Intelligence. Cambridge, MA: Harvard Business Review Press.

15 Goldfarb, Avi, and Daniel Trefler. 2018. "AI and International Trade." NBER Working Paper 24254, National Bureau of Economic Research, Cambridge, MA.

16 For an overview of the compendium of articles, see: Goldfarb, A., Gans, J., & Agrawal, A. (Eds.) (2019). The

Economics of Artificial Intelligence: An Agenda. University of Chicago Press. A list of the articles can be accessed via https://www.nber.org/books/agra-1

17 Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz Harvard, B., Lyons, T., Manyika, J., Carlos Niebles, J., & Mishra, S. (2019). Artificial Intelligence Index 2019 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, p. 89. https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

18 The list also includes two Korean companies but no Indian companies. See WIPO. (2019). WIPO Technology Trends 2019: Artificial Intelligence. In Geneva: World Intellectual Property Organization. World Intellectual Property Organization, p. 60. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf

19 Pinto, A. R. (2018). Digital Sovereignty or Digital Colonialism? SUR 27, 15(27). https://sur.conectas.org/wp-content/uploads/2018/07/sur-27-ingles-renata-avila-pinto.pdf

20 Nick Couldry and Ulises A. Mejias (2019) The Costs of Connection. Stanford: Stanford University Press.

21 Birhane, A. (2020). Algorithmic Colonization of Africa. Scripted, 17(2), 389–409. https://doi.org/10.2966/scrip.170220.389

22 Ray, T. (2020). The quest for cyber sovereignty is dark and full of terrors. Observer Research Foundation. https://www.orfonline.org/expert-speak/the-quest-for-cyber-sovereignty-is-dark-and-full-of-terrors-66676/

23 For an overview of the overall data base, see: Dalle, J., M. den Besten and C. Menon (2017), "Using Crunchbase for economic and managerial research," OECD Science, Technology and Industry Working Papers, No. 2017/08, OECD Publishing, Paris. https://doi.org/10.1787/6c418d60-en

24 Lopez Gonzales, J., & Casalini, F. (2018). Trade and cross-border data flows. OECD Working Party of the Trade Committee. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=TAD/TC/WP(2018)19/FINAL&docLanguage=En

25 In fact, the GDPR's predecessor, the Directive 95/46/EC (DPD), had very similar provisions on cross-border transfer. However, the level of harmonization has certainly extended since the GDPR's adoption. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

26 Yakovleva, Svetlana, and Kristina Irion. 2020. "Pitching Trade against Privacy: Reconciling EU Governance of Personal Data Flows with External Trade." International Data Privacy Law 0(0): 1–21. https://doi.org/10.1093/idpl/ipaa003

27 Velli, Federica. 2019. "The Issue of Data Protection in EU Trade Commitments: Cross-Border Data Transfers in GATS and Bilateral Free Trade Agreements." European Papers http://www.europeanpapers.eu/en/europeanforum/issue-of-data-protection-in-eu-trade-commitments#_ftnref10

28 Velli, Federica. 2019. "The Issue of Data Protection in EU Trade Commitments: Cross-Border Data Transfers in GATS and Bilateral Free Trade Agreements." European Papers 4(3): 881–94. http://www.europeanpapers.eu/en/system/files/pdf_version/EP_EF_2019_I_022_Federica_Velli_0.pdf

29 https://www.thehindubusinessline.com/info-tech/EU-not-ready-to-give-India-%E2%80%98data-secure%E2%80%99-status/article20624400.ece

30 BCRs must include a mechanism to make the BCRs legally binding on group companies. Among other things, the BCRs must: (1) specify the purposes of the transfer and affected categories of data; (2) reflect the requirements of the GDPR; (3) confirm that the EU- based data exporters accept liability on behalf of the entire group; (4) explain complaint procedures; and (5) provide mechanisms for ensuring compliance (e.g., audits).

31 Lopez Gonzales, Javier, and Francesca Casalini. 2018. Trade and Cross-Border Data Flows. Paris: OECD Working Party of the Trade Committee, pp. 24-26. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=TAD/TC/WP(2018)19/FINAL&docLanguage=En

32 European Commission. 2020. Communication COM(2020) 66 Final on "A European Strategy for Data." Brussels: European Commission, p. 9. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

33 BMWi, and BMBF. 2019. Project GAIA-X: A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem. Berlin: Bundesministerium für Wirtschaft und Energie (BMWi), pp. 7–8. https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.pdf?__blob=publicationFile&v=4

34 Manancourt, Vincent. 2020. "Europe's Data Grab." Politico (online). https://www.politico.eu/article/europe-data-grab-protection-privacy/

35 Biancotti, C. (2019). India's ill-advised pursuit of data localization. https://www.piie.com/blogs/realtime-economic-issues-watch/indias-ill-advised-pursuit-data-localization

36 European Commission. (2017). Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, 10.01.2017. European Commission, p. 7ff. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017SC0002&from=EN

37 Gartner. (2019). Market Share Analysis: IaaS and IUS, Worldwide, 2018. https://www.gartner.com/en/newsroom/press-releases/2019-07-29-gartner-says-worldwide-iaas-public-cloud-services-market-grew-31point3-percent-in-2018

38 https://www.ft.com/content/e489abba-0dc5-11e8-8eb7-42f857ea9f09, https://www.politico.eu/article/europe-trade-data-protection-privacy/

39 Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence 1(9): 389–99. https://www.nature.com/articles/s42256-019-0088-2

40 AI HLEG. 2019. Ethics Guidelines for Trustworthy AI. Brussels: European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

41 VDE, and Bertelsmann Stiftung. 2020. Journal of Corporate Citizenship: From Principles to Practice – An Interdisciplinary Framework to Operationalise AI Ethics. Gütersloh: Bertelsmann Stiftung. https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf

42 European Commission. (2018). Artificial Intelligence for Europe, COM(2018) 273 final, 25.04.2018. European Commission, p. 3. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625

43 McKinsey Global Institute. (2019). "Notes from the AI frontier tackling Europe's gap in digital and AI," p. 10. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20europes%20gap%20in%20digital%20and%20ai/mgi-tackling-europes-gap-in-digital-and-ai-feb-2019-vf.ashx

44 Deloitte. (2017). Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. https://doi.org/10.2759/781960, p. 16. Cited in Arnaut, C., Pont, M., Scaria, E., Berghmans, A., & Leconte, S. (2018). Data sharing between companies in Europe. https://doi.org/10.2759/354943

45 Heumann, S., & Jentzsch, N. (2019). Think Tank für die Gesellschaft im technologischen Wandel Wettbewerb um Daten Über Datenpools zu Innovationen. Stiftung Neue Verantwortung. https://www.stiftung-nv.de/sites/default/files/wettbewerb_um_daten.pdf

46 VDMA, Roland Berger, & Deutsche Messe. (2018). Plattformökonomie im Maschinenbau. https://www.vdma.org/documents/15012668/26471342/RB_PUB_18_009_VDMA_Plattformökonomie-06_1530513808561.pdf/f4412be3-e5ba-e549-7251-43ee17ec29d3

47 For three case studies, see: Micheletti, G., Carnelley, P., & Spinoni, E. (2019). How Big Data is driving AI: Selected Examples of AI Applications across European Industries – Update of the European Data Market SMART 2016/0063. http://datalandscape.eu/sites/default/files/report/D3.4_Big_Data_and_AI_15.03.2019.pdf

48 Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. Prediction Machines: The Simple Economics of Artificial Intelligence. Cambridge, MA: Harvard Business Review Press.

49 Singh, R., Kalra, M. K., Nitiwarangkul, C., Patti, J. A., Homayounieh, F., Padole, A., Rao, P., Putha, P., Muse, V. V., Sharma, A., & Digumarthy, S. R. (2018). Deep learning in chest radiography: Detection of findings and presence of change. PLoS ONE, 13(10). https://doi.org/10.1371/journal.pone.0204155

50 Platforms such as Kaggle and some firms (e.g. Google research) are sharing AI-relevant datasets online. An older list (2017) of dozens of such datasets was compiled by the Indian image recognition firm KritiKal. https://kritikalsolutions.com/685-outstanding-free-data-sources-for-2017/

51 https://factordaily.com/indian-ai-has-a-dataset-problem/

52 For NLP examples, see: Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), pp. 183–186. https://doi.org/10.1126/science.aal4230, and Jentzsch, S., Schramowski, P., Rothkopf, C., & Kersting, K. (2019). The Moral Choice Machine: Semantics Derived Automatically from Language Corpora Contain Human-like Moral Choices. Proceedings of AIES, (68). https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19_paper_68.pdf

53 https://uidai.gov.in/aadhaar_dashboard/

54 Aadhaar is a 12-digit unique identity number that can be obtained by residents of India, based on their biometric and demographic data. The official website is https://www.uidai.gov.in

55 https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

56 Ravi, S. (2018). Is India ready to JAM? (Issue August). Brookings Institution India. https://www.brookings.edu/wp-content/uploads/2018/08/JAM-paper.pdf

57 "What is India Stack?" https://www.indiastack.org/about/. Last accessed 19.02.2020.

58 Olivier Garret. (2017). India Is Likely to Become the First Digital, Cashless Society. Forbes (Online). https://www.forbes.com/sites/oliviergarret/2017/06/28/india-is-likely-to-become-the-first-digital-cashless-society/#7c68372d3c80

59 "Why Aadhaar Paperless Offline e-KYC?" https://uidai.gov.in/2-uncategorised/11320-aadhaar-paperless-offline-e-kyc-3.html. Last accessed 19.02.2020.

60 PwC. (2019). Emerging technologies disrupting the financial sector. PwC. https://www.pwc.in/assets/pdfs/consulting/financial-services/fintech/publications/emerging-technologies-disrupting-the-financial-sector.pdf

61 Ernst & Young. (2019). Global FinTech Adoption Index 2019, See figure 1. https://www.ey.com/en_gl/ey-global-fintech-adoption-index

62 PwC. (2019). Emerging technologies disrupting the financial sector. PwC. https://www.pwc.in/assets/pdfs/consulting/financial-services/fintech/publications/emerging-technologies-disrupting-the-financial-sector.pdf. For Germany: Statista. https://www.statista.com/outlook/295/137/fintech/germany

63 Anandi Chandrashekhar. (2020). Google Pay India learnings will be taken to global markets: Alphabet CEO Sundar Pichai. The Economic Times (Online). https://economictimes.indiatimes.com/tech/internet/google-pay-india-learnings-will-be-taken-to-global-markets-alphabet-ceo-sundar-pichai/articleshow/73934899.cms

64 OECD. (2019). OECD Economic Surveys: India – December 2019. OECD Publishing. http://www.oecd.org/economy/surveys/India-2019-OECD-economic-survey-overview.pdf

65 McKinsey Global Institute. (2019). Digital India – Technology to transform a connected nation, p. 28. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-india-technology-to-transform-a-connected-nation

66 GSMA. (2019). Connected Society – The State of Mobile Internet Connectivity 2019. https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-State-of-Mobile-Internet-Connectivity-Report-2019.pdf

67 Deutsche Bank Research. (2019). Imagine 2030, p. 32. https://www.dbresearch.de/PROD/RPS_DE-PROD/PROD0000000000503196/Imagine_2030.PDF

68 Competition Commission of India. (2020). "Market Study on E-Commerce in India," p. 6. https://www.cci.gov.in/sites/default/files/whats_newdocument/Market-study-on-e-Commerce-in-India.pdf

69 Others declared that the digital economy in India has the potential to reach $1 trillion by 2022 – out of a projected GDP of $3.85 trillion in 2022. See: Ernst & Young. (2019). Propelling India to a trillion dollar digital economy - Implementation roadmap to NDCP 2018. ASSOCHAM India & EY. https://www.meity.gov.in/writereaddata/files/india_trillion-dollar_digital_opportunity.pdf

70 McKinsey Global Institute. (2019). Digital India – Technology to transform a connected nation. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-india-technology-to-transform-a-connected-nation

71 Naik, S. (2019). IndiGen project – How mapping of genomes could transform India's healthcare. ThePrint (Online), 29 November. https://theprint.in/science/indigen-project-how-mapping-of-genomes-could-transform-indias-healthcare/327670/

72 Mehrotra, S. (2020). "Make in India": The Components of a Manufacturing Strategy for India. The Indian Journal of Labour Economics. https://doi.org/10.1007/s41027-019-00201-9

73 Economist Intelligence Unit. 2018. Who Is Ready for the Coming Wave of Automation? London: EIU. https://www.automationreadiness.eiu.com/static/download/PDF.pdf

74 Holtkamp, Bernhard, and Anandi Iyer. 2018. Industry 4.0 – The Future of Indo-German Industrial Collaboration. Gütersloh: Bertelsmann Stiftung, p. 21 https://www.bertelsmann-stiftung.de/en/publications/publication/did/industry-40

75 Prasad, S., Raghavan, M. Chugh, B. Singh, A. (October 2019). "Implementing the Personal Data Protection Bill: Mapping Points of Action for Central Government and the Future Data Protection Authority in India." Dvara Research. https://www.dvara.com/research/wp-content/uploads/2019/10/Policy-Brief-Implementing-the-Personal-Data-Protection-Bill-in-India.pdf

76 The following represents a selection of points relevant to our topic. They draw heavily on Nishith Desai Associates. (2019). Analysis of the New Data Protection Law Proposed in India. http://www.nishithdesai.com/fileadmin/user_upload/pdfs/NDA%20Hotline/Analysis_of_the_new_Data_Protection_Law_Dec2419.pdf

77 See Section 33 (2).

78 PDP considers three categories of data: Sensitive data includes information on financials, health, sexual orientation, genetics, transgender status, caste, and religious belief. Critical data includes information that the government stipulates from time to time as extraordinarily important, such as military or national security data. The third is a

general category, which is not defined but contains the remaining data. DPB prescribes specific requirements that data fiduciaries must follow for the storage and processing for each data class. Described in Govindarajan, V., Srivastava, A., & Enache, L. (2019). How India Plans to Protect Consumer Data. Harvard Business Review (Online). https://hbr.org/2019/12/how-india-plans-to-protect-consumer-data

79 https://www.livemint.com/opinion/columns/opinion-we-need-a-cost-benefit-analysis-of-data-localization-1564513414995.html

80 https://www.thehindubusinessline.com/info-tech/personal-data-protection-bill-compliance-costs-to-rise-for-india-inc/article30971116.ece

81 Economic Times, March 6, 2020, https://economictimes.indiatimes.com/tech/internet/data-bill-global-trade-bodies-raise-privacy-concerns/articleshow/74503958.cms

82 NITI Aayog. (2018). National Strategy for Artificial Intelligence – #AIforALL. https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf

83 Groth, O. J., Nitzberg, M., & Zehr, D. (2019). Comparison of National Strategies to Promote Artificial Intelligence – Part 2. Konrad-Adenauer-Stiftung.

84 NITI Aayog. (2018). National Strategy for Artificial Intelligence – #AIforALL, p. 8. https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf

85 Other landmark documents that talk about ethical aspects of privacy and data protection include the Charter of Fundamental Rights of the European Union (Article 1 states "human dignity" as inviolable and that it should be protected at all time; the charter also talks about the notion of fairness in data processing) and the Privacy Principles of the OECD from 1980 (the principles refer to fair and non-excessive data collection; the GDPR incorporates traditional data protection principles established in this document).

86 Hijmans, Hielke and Raab, Charles D., Ethical Dimensions of the GDPR (July 30, 2018). https://ssrn.com/abstract=3222677

87 https://www.linklaters.com/en/insights/data-protected/data-protected---india

88 The Information Technology Act 2000. https://indiacode.nic.in/bitstream/123456789/1999/3/A2000-21.pdf. It is India's only cyber law and was largely enacted to provide a legal infrastructure for electronic commerce in the nation. With the increasing trend towards electronic activities, the passage of this Act was considered to be of paramount importance. The Act also talks about cybercrimes, cyberterrorism and the Adjudicator and the Cyber Appellate Tribunal. Drafted by the Ministry of Communication and IT (now known as MEITY).

89 This consent can also be withdrawn in writing.

90 Rule 6 of IT Rules 2011. https://www.prsindia.org/uploads/media/IT%20Rules/IT%20Rules%202011.pdf

91 Section 7 IT Rules: Bodies corporate can transfer sensitive personal data to any other body corporate or person within or outside India, provided the transferee ensures the same level of data protection that the body corporate maintained, as required by the IT Rules. A data transfer is only allowed if it is required for the performance of a lawful contract between the data controller and the data subjects; or the data subjects have consented to the transfer.

92 It states that if a body corporate possesses, deals with or handles any sensitive personal data or information and negligently does not put in place safety practices that are reasonable (as elaborated under Rule 8 of the IT Rules, 2011), and if the same causes wrongful loss or wrongful gain to any person, then the body corporate is liable to pay compensation to the person to whom the data belongs.

93 72A Punishment for disclosure of information in breach of lawful contract. -Save as otherwise provided in this Act or any other law for the time being in force, any person including an intermediary who, while providing services under the terms of lawful contract, has secured access to any material containing personal information about another person, with the intent to cause or knowing that he is likely to cause wrongful loss or wrongful gain discloses, without the consent of the person concerned, or in breach of a lawful contract, such material to any other person, shall be punished with imprisonment for a term which may extend to three years, or with fine which may extend to five lakh rupees, or with both (https://indiankanoon.org/doc/69360334/)

94 However, for the same to be applicable, the said contravention or offence must involve a computer, computer system or computer network located in India.

95 Subramanium Aditi, Das Sujan. (2019). https://thelawreviews.co.uk/edition/the-privacy-data-protection-and-cybersecurity-law-review-edition-6/1210048/india

96 https://www.linklaters.com/en/insights/data-protected/data-protected---india

97 It applies to individuals and bodies (including a company, a firm, sole proprietorship or other association of individuals engaged in professional or commercial activities), but not the government.

98 https://www.tribuneindia.com/news/archive/nation/rs-500-10-minutes-and-you-have-access-to-billion-aadhaar-details-523361. Two other data breaches were also reported, including granular details of 6.7 million people who have received subsidies and that of 166,000 government workers in the state of Jharkand https://www.deccanherald.com/national/a-look-at-data-breaches-cyberattacks-india-saw-in-2019-785987.html

99 https://economictimes.indiatimes.com/tech/internet/german-firm-finds-one-million-files-of-indian-patients-leaked/articleshow/73921423.cms?from=mdr

100 This most likely represents an upper boundary because Crunchbase does not offer a comprehensive list of Indian firms and is biased towards younger and tech-related firms. The database contains about 30,000 Indian companies. The total number of active registered companies according to the Economic Census is more than one million. Thus, the "AI penetration" could also be read as "per 1,000 manufacturing firms headquartered in India and listed in Crunchbase."

101 Gartner. (2019). "2019 CIO Survey." https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have

102 Barclays and MMC Ventures (2019). "State of AI 2019". https://www.stateofai2019.com/introduction/

103 https://www.livemint.com/companies/start-ups/as-demand-soars-zomato-swiggy-turn-to-ai-machine-learning-to-drive-growth-11572285380613.html

104 https://www.tcs.com/cargo-tech-iot-platform-for-digital-services

105 https://www.infosys.com/services/ai-automation/insights/consulting-driven-approach.html

106 Barclays and MMC Ventures (2019). State of AI 2019. https://www.stateofai2019.com/introduction/

107 https://economictimes.indiatimes.com/tech/ites/it-services-tech-startups-put-india-on-route-to-ai-etilc/articleshow/72324818.cms?from=mdr

108 https://www.alliedmarketresearch.com/artificial-intelligence-in-healthcare-market

109 69.3% of current health expenditure in 2017, WHO. (2017). "Health Financing Profile 2017" https://apps.who.int/iris/bitstream/handle/10665/259642/HFP-IND.pdf?sequence=1&isAllowed=y

110 In 2018. https://www.ibef.org/industry/healthcare-India.aspx

111 Bengal Chamber of Commerce and PWC. (2018). "Reimagining the possible in the Indian healthcare ecosystem with emerging technologies." https://www.pwc.in/industries/healthcare/reimagining-the-possible-in-the-indian-healthcare-ecosystem-with-emerging-technologies.html

112 https://health.economictimes.indiatimes.com/news/health-it/india-bullish-on-ai-in-healthcare-without-ehr/73118990

113 Census 2011 India, compiled by authors. http://pibmumbai.gov.in/scripts/detail.asp?releaseId=E2011IS3

114 Bengal Chamber of Commerce and PWC. (2018). "Reimagining the possible in the Indian healthcare ecosystem with emerging technologies." https://www.pwc.in/assets/pdfs/publications/2018/reimagining-the-possible-in-the-indian-healthcare-ecosystem-with-emerging-technologies.pdf

115 Indian Brand Equity Foundation. (2019). "Healthcare". https://www.ibef.org/download/healthcare-jan-2019.pdf

116 INR ~431.97 Bn. conversion rate 64 Research and Markets. (2019). "Artificial Intelligence (AI) in Healthcare Market in India (2018–2023)" https://www.researchandmarkets.com/reports/4787271/artificial-intelligence-ai-in-healthcare-market?utm_source=BW&utm_medium=PressRelease&utm_code=xh8qm2&utm_campaign=1279677+-+Artificial+Intelligence+(AI)+in+Healthcare+Market+in+India+is+Expected+to+Grow+at+a+Rate+of+40%25+by+2021%2c+Impacting+the+Doctor-Patient+Ratio+During+the+Forecast+Period%2c+2018-2023&utm_exec=anwr281prd

117 https://inc42.com/features/indian-startup-ecosystem-sectors/

118 Indian Brand Equity Foundation. (2020). "Healthcare Industry in India." Updated on March 2020. https://www.ibef.org/industry/healthcare-india.aspx

119 https://www.siemens-healthineers.com/en-in/digital-health-solutions/value-themes/artificial-intelligence-in-healthcare#Our_solutions

120 https://www.philips.co.in/healthcare/resources/feature-detail/ultrasound-heartmodel

121 https://www.livemint.com/AI/zAlE15t4dKOhHYFY-Wnh5MJ/Reinventing-the-healthcare-sector-with-Artificial-Intelligen.html

122 https://www.wired.com/2017/06/googles-ai-eye-doctor-gets-ready-go-work-india/

123 Gulshan Varun, et al. (2016). "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45732.pdf

124 https://www.medicalbuyer.co.in/indian-healthcare-is-all-set-to-be-transformed-by-ai/

125 Phillips Mark, et al. ()2020). "Genomics: data sharing needs an international code of conduct." https://www.nature.com/articles/d41586-020-00082-9?utm_source=Nature+Briefing&utm_campaign=1ea97fb35e-briefing-dy-20200206&utm_medium=email&utm_term=0_c9dfd39373-1ea97fb35e-44332685

126 BSA. (2017). Cross-border data flows. https://www.bsa.org/files/policy-filings/BSA_2017CrossBorderDataFlows.pdf
127 NITI Aayog. (2018). "National Health Stack: Strategy and approach." https://niti.gov.in/writereaddata/files/document_publication/NHS-Strategy-and-Approach-Document-for-consultation.pdf
128 https://www.medicalbuyer.co.in/indian-healthcare-is-all-set-to-be-transformed-by-ai/
129 To circumvent this, they often need to identify and partner with doctors and hospitals.
130 Ministry of External Affairs, India. (2019). Joint Statement during the visit of Chancellor of Germany to India. https://mea.gov.in/bilateral-documents.htm?dtl/31991/Joint+Statement+during+the+visit+of+Chancellor+of+Germany+to+India
131 https://www.medicalbuyer.co.in/indian-healthcare-is-all-set-to-be-transformed-by-ai/
132 See for example, Abernathy, F. H., Dunlop, J. T., Hammond, J. H., & Weil, D. (2000). Retailing and supply chains in the information age. Technology in Society, 22(1), 5–31. The authors talk about how retail stores have dramatically increased their adoption of information technology since the 1980s and now use it for distribution, forecasts and plan production and managing their supplier and customer relations. doi:10.1016/s0160-791x(99)00039-1)
133 https://www.emarketer.com/content/global-ecommerce-2019
134 https://www.retaildive.com/news/retail-spending-on-ai-to-reach-73b-by-2022/516170/
135 Juniper Research. "How AI can revive retail." https://www.juniperresearch.com/document-library/white-papers/how-ai-can-revive-retail?utm_campaign=aiinretail18pr1&utm_source=businesswire&utm_medium=email
136 UNCTAD (2019). "Digital economy 2019 report." https://unctad.org/en/PublicationsLibrary/der2019_en.pdf
137 https://www.emarketer.com/content/global-ecommerce-2019
138 https://www.ibef.org/industry/ecommerce.aspx
139 KPMG (2018). "Start-up ecosystem in India – Growing or matured?" https://assets.kpmg/content/dam/kpmg/in/pdf/2019/01/startup-landscape-ecosystem-growing-mature.pdf
140 https://iapp.org/news/a/cross-border-data-transfers-in-privacy-driven-india/
141 https://www.thehindu.com/business/Industry/flipkart-uses-ai-to-tackle-address-frauds-complexities/article23467831.ece
142 Redseer (2018). "E-tailing in India". https://redseer.com/wp-content/uploads/2018/04/e-tailing-in-india-redseer-perspective-2018.pdf
143 Xiao Han, Rasul Kashif, Vollgraf Roland. (2017). "Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms." https://arxiv.org/pdf/1708.07747.pdf
144 Several of these databases can be found online, selling contact information and demographic data of Indian shoppers across income groups and professions. For example, the E-commerce database (http://www.ecommercedatabase.in/Buy-Ecommerce-Email-Data-Database.html) and Aeroleads (https://aeroleads.com/database), Bol7 (https://www.bol7.com/mobile-number-database/).
145 https://www.thehindubusinessline.com/opinion/the-dirty-world-of-data-brokering/article23324882.ece
146 Omidiyar Network and Deloitte (2019). "Unlocking the Potential of India's Data Economy: Practices, Privacy and Governance." https://www.omidyarnetwork.in/insights/unlocking-the-potential-of-indias-data-economy-practices-privacy-and-governance
147 https://iapp.org/news/a/cross-border-data-transfers-in-privacy-driven-india/
148 https://economictimes.indiatimes.com/industry/services/retail/draft-e-commerce-policy-good-intentions-but-no-solutions/articleshow/68145183.cms
149 Competition Commission of India. (2020). Market study on e-commerce in India. https://www.cci.gov.in/sites/default/files/whats_newdocument/Market-study-on-e-Commerce-in-India.pdf
150 https://www.livemint.com/companies/news/cci-orders-antitrust-probe-against-amazon-flipkart-11578921450124.html
151 In their analysis of the German AI Strategy, Groth and Straube wrote in their headline: "A late but multi-faceted approach for 'AI made in Germany'." See Groth, O. J., & Straube, T. (2019). Evaluation of the German AI Strategy – Part 3. Konrad-Adenauer-Stiftung. https://www.kas.de/documents/252038/4521287/Evaluation+of+the+German+AI+Strategy+Part+3.pdf/a1069da4-02f9-56ba-babc-6c85627a7a39?version=1.0&t=1562322073583
152 Oxford Insights. (2019). Government Artificial Intelligence Readiness Index 2019. https://ai4d.ai/wp-content/uploads/2019/05/ai-gov-readiness-report_v08.pdf
153 Groth, O. J., Nitzberg, M., & Zehr, D. (2019). Comparison of National Strategies to Promote Artificial Intelligence. Konrad-Adenauer-Stiftung.

154 Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz Harvard, B., Lyons, T., Manyika, J., Carlos Niebles, J., & Mishra, S. (2019). Artificial Intelligence Index 2019 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

155 Data from stat.oecd.org, retrieved 19.03.2020. Categorization of "potentially ICT-enabled services" from Grimm (2016). Analysis by CPC Analytics. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=STD/CSSP/WPTGS/RD(2017)2&docLanguage=En

156 ITU. (2018). Assessing the Economic Impact of Artificial Intelligence. International Telecommunication Union, p. 16. https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf

157 For Germany: BMWi, & ZEW. (2018). Digital Economy Monitoring Report 2018. Bundesministerium für Wirtschaft und Energie (BMWi). http://ftp.zew.de/pub/zew-docs/gutachten/ZEWKantarTNS_Monitoring_Report_ICT_2018_Compact.pdf. For India: MeitY. (2019). India's trillion-dollar digital opportunity. Ministry of Electronics & Information Technology, p. 6. https://meity.gov.in/writereaddata/files/india_trillion-dollar_digital_opportunity.pdf

158 Ministry of Foreign Affairs India. (2019). Joint Statement during the visit of Chancellor of Germany to India. https://mea.gov.in/bilateral-documents.htm?dtl/31991/Joint+Statement+during+the+visit+of+Chancellor+of+Germany+to+India

159 Ibid.

160 Merkel, A. (2020). Speech by Federal Chancellor Dr Angela Merkel at the 63rd Annual General Meeting of the Indo-German Chamber of Commerce in New Delhi on 2 November 2019. New Delhi. https://www.bundeskanzlerin.de/bkin-en/news/speech-by-federal-chancellor-dr-angela-merkel-at-the-63rd-annual-general-meeting-of-the-indo-german-chamber-of-commerce-in-new-delhi-on-2-november-2019-1704858

161 Measured here as the share of imports in the following sectors: Insurance and pension services; financial services; charges for intellectual property use not included elsewhere; telecommunications, computer and information services; audiovisual and related services.

162 OECD. (2019). Going Digital: Shaping Policies, Improving Lives. OECD Publishing: Paris. https://www.oecd-ilibrary.org/sites/c62c199f-en/index.html?itemId=/content/component/c62c199f-en, p. 135 [figure 8.2].

163 RBI. (2020). Foreign Direct Investment Flows to India: Country-wise and Industry-wise. Updated figures from Aug 29, 2019. Reserve Bank of India. https://m.rbi.org.in/Scripts/AnnualReportPublications.aspx?Id=1278

164 Bertelsmann Stiftung, Ernst and Young, & Confederation of Indian Industry (CII). (2017). Indian Investments in Germany. Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/fileadmin/files/user_upload/20170922_Indian_investments_in_Germany_EN.pdf

165 Ministry of Foreign Affairs India. (2019). Joint Statement during the visit of Chancellor of Germany to India. https://mea.gov.in/bilateral-documents.htm?dtl/31991/Joint+Statement+during+the+visit+of+Chancellor+of+Germany+to+India

166 https://www.igstc.org/

167 Bundesbericht Forschung und Innovation. (2018). http://dip21.bundestag.de/dip21/btd/19/026/1902600.pdf

168 IGSTC. (2019). Newsletter of IGSTC May-August 2019. https://www.igstc.org/images/newsletters/157950759820200120.pdf

169 https://www.igcs-chennai.org/

170 Embassy of India in Berlin. (2020). Joint Statement during the visit of Chancellor of Germany to India. https://mea.gov.in/bilateral-documents.htm?dtl/31991/Joint+Statement+during+the+visit+of+Chancellor+of+Germany+to+India

171 ODI. (2020). How new data institutions could be the key to radically improving our health and wellbeing. https://theodi.org/article/how-new-data-institutions-could-be-the-key-to-radically-improving-our-health-and-wellbeing/

172 In general terms, "[T]rusts are legal instruments that appoint a steward (trustee) to manage an asset for a purpose — such as conservation of land or maximizing value — on behalf of a beneficiary or beneficiaries who own the asset. Data trusts are legal trusts that manage data, or the rights to data." McDonald, S. (2019). Reclaiming Data Trusts. https://www.cigionline.org/articles/reclaiming-data-trusts

173 McDonald, S. (2019). Reclaiming Data Trusts. https://www.cigionline.org/articles/reclaiming-data-trusts

174 ODI. (2019). INSIGHT: exploring access to eye health data. https://theodi.org/project/insight/

175 See: BMWi, BMBF. (2019) Project GAIA-X: A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem. Berlin: Bundesministerium für Wirtschaft und Energie (BMWi), https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.pdf?__blob=publicationFile&v=4 and Dureau (2018): Dezentrales Machine Learning – lernen mit verschlüsselten Nutzerdaten. https://www.bigdata-insider.de/dezentrales-machine-learning-lernen-mit-verschluesselten-nutzerdaten-a-784058/

176 OECD. (2019). Measuring the Digital Transformation – A roadmap for the future. In Measuring the Digital Transformation. OECD Publishing.

177 "A definition of AI: Main capabilities and scientific disciplines," AI HLEG, 18 December 2018 https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december.pdf. An alternative definition of AI Systems is offered by the OECD Expert Group on AI: An AI system is "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. It uses machine and/or human-based inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with ML [machine learning] or manually); and use model inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy." See: Proposal by the Expert Group on Artificial Intelligence at the OECD (AIGO), OECD, Paris. In: OECD. (2019). Artificial Intelligence in Society. Paris: OECD Publishing, p. 15. https://doi.org/10.1787/eedfee77-en

178 Ibid., p. 27.

179 Ibid., p. 28.

180 Accenture. (2019). Maximize collaboration through secure data sharing. https://www.accenture.com/_acnmedia/PDF-114/Accenture-PPC-Techniques.PDF#zoom=40

We are grateful for the valuable contributions and insights provided by our interview partners as well as the participants in our digital roundtable on the topic in July 2020.

Publishing Information

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Germany
Phone +49 5241 81-0
www.bertelsmann-stiftung.de

Executive Editors
Dr. Maximilian von Laer
Bernhard Bartsch

Authors
Christian Franz,
Sanjana Krishnan
Sahil Deo
(CPC Analytics)
Kashyap Raibagi, research assistance

Editing
Tim Schroder, Frankfurt

Design
Nicole Meyerholz, Bielefeld

Cover
© monsitj – stock.adobe.com

Printing
Lensing Druck, Dortmund

DOI 10.11586/2020065

Address | Contact

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh | Germany
Phone +49 5241 81-0

Dr. Maximilian von Laer
Project Manager
Program Germany and Asia
Phone +49 5241 81-81836
maximilian.vonlaer@bertelsmann-stiftung.de

Bernhard Bartsch
Senior Expert China and Asia Pacific
Program Germany and Asia
Phone +49 30 275788-129
bernhard.bartsch@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de

BertelsmannStiftung