Opportunities and Challenges for Artificial Intelligence in India

Shivaram Kalyanakrishnan
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
shivaram@cse.iitb.ac.in

Sarayu Natarajan
King's India Institute
King's College London
sarayu.natarajan@kcl.ac.uk

Rahul Alex Panicker
Embrace Innovations
rahulap@alumni.stanford.edu

Shreya Rao
Rao Law Chambers and
Azim Premji University
shreya@raolawchambers.com

ABSTRACT

In the future of India lies the future of a sixth of the world's population. As the Artificial Intelligence (AI) revolution sweeps through societies and enters daily life, its role in shaping India's development and growth is bound to be substantial. For India, AI holds promise as a catalyst to accelerate progress, while providing mechanisms to leapfrog traditional hurdles such as poor infrastructure and bureaucracy. At the same time, an investment in AI is accompanied by risk factors with long-term implications on society: it is imperative that risks be vetted at this early stage. In this paper, we describe opportunities and challenges for AI in India. We detail opportunities that are cross-cutting (bridging India's linguistic divisions, mining public data), and also specific to one particular sector (healthcare). We list challenges that originate from existing social conditions (such as equations of caste and gender). Thereafter we distill out concrete steps and safeguards, which we believe are necessary for robust and inclusive development as India enters the AI era.

CCS CONCEPTS

• **Computing methodologies** → *Artificial intelligence*; • **Social and professional topics** → Government technology policy;

KEYWORDS

Artificial Intelligence, India

1 INTRODUCTION

Investigations into the effect of technology on society are often structured "vertically" around topics such as ethics [15], law [12], economic productivity and employment [11], and the implications of social markers such as gender [59] and race [16]. In this paper, we adopt a "horizontal" framing that views the totality of such questions from a fixed perspective: the role of Artificial Intelligence (AI) in the ongoing development of India. Our choice of a country-specific perspective is not new. For example, Little (1999) examines the effects of the global production system on several East Asian countries; the recently undertaken *AI100* study [57] considers a variety of domains at the intersection of AI and a "typical North American city". A paucity of academic literature on the implications of AI for India motivates a unified treatment of relevant technical and non-technical questions. Our paper aims to provide a framework to which technologists, social scientists, and policy makers can all contribute.

In India's future lies the future of a sixth of the world's population—enough reason by itself to track the country's tryst with AI. Of equal interest is India's unique social, cultural, economic, and political context, which has the potential to magnify both the benefits and the risks of AI. With a large, young workforce [62], a fast-growing economy [41], and a vibrant, resilient democracy [48], India presents an opportunity for AI applications to have tremendous reach and scale (and helping create abundance). AI-driven interventions can enhance public services: for example, streamlining the public distribution system, and reducing the costs of law enforcement. AI can also enhance private services, such as the use of AI-enabled personalised healthcare, or robots in production lines. On the other hand, India's *challenges*—varying from income inequality [1, 17] and caste-based discrimination [4] to linguistic diversity [42]—are also magnified by the size and variety of the population. Yet other societal challenges, such as malnutrition [39] and girls' education [43], may best be tackled through means unrelated to AI. Where AI can indeed make meaningful contributions, its solutions will often have to withstand cultural forces shaped by millennia of civilisational history.

Since India is significantly behind many other countries in its technological development, it is natural for technologists and policy makers to look to transplant successful ideas from other contexts into India. A growing body of literature warns of the inefficiency, even danger, of such an approach [9, 37, 50]. The main thesis of this paper is indeed the need to plan "AI for India" from the *bottom*

up, by paying attention to India's social, political, cultural, and economic configuration.

We present our thesis by first outlining a variety of technical problems that arise in India's unique context: in its linguistic diversity, legacy public records, and healthcare system. We hope that this compendium of illustrative problems, provided in Section 2, will enthuse and enable technologists to work on socially-relevant challenges. The benefits could be substantial. Equally, the advent of AI could bring with it a variety of risks. In Section 3, we specifically highlight existing gaps in Indian society (for example, based on caste and gender) that AI-driven development could widen. In Section 4, we consolidate our discussion and propose concrete steps and safeguards for carrying forward AI in India. We present our conclusions in Section 5.

We are not aware of literature on AI and India that is similar in scope to our paper. Our focus is broader when compared to summaries of the state of AI *research* in different countries: for example, Israel [21], Singapore [60], and India itself [28]. Closest in spirit to our paper is one by Vempati (2016), which is intended as a "wake-up call" to Indian policy makers. Vempati presents geopolitical considerations, including comparisons with China, as he prompts urgency in adopting AI. We describe a complementary viewpoint that primarily looks *inside* India: in so doing, we find both possibilities and pitfalls, which we describe in some detail.

2 OPPORTUNITIES FOR AI-DRIVEN DEVELOPMENT

For India, AI holds promise as a catalyst to accelerate progress and to leapfrog traditional hurdles such as poor infrastructure and bureaucracy. In nearly every sector—finance, healthcare, law enforcement, transportation, agriculture, environmental conservation—one finds applications in which AI can be effective. In a timely move, the Indian government has recently constituted a task force precisely to identify openings for AI across sectors and guide policy [40]. In this section we describe some uniquely (at any rate, typically) Indian problems and their amenability to AI. Rather than enumerate a long list, we restrict our focus to three illustrative problems, which we present in some detail. The first two (in sections 2.1 and 2.2) cut across different sectors; the third (Section 2.3) relates to a particular sector.

2.1 Scaling up NLP/ASR for Indian Languages

Over 700 languages are spoken in India, making it among the most multilingual of countries [42]. The languages span six *families*: Indo-Aryan, Dravidian, Austroasiatic, Sino-Tibetan, Tai–Kadai, and Great Andamanese. At least 20 languages are first languages to over a million speakers. Since a large section of the population is either monolingual or bilingual, language naturally becomes a barrier to communication and access to information. Written in English, this very paper is inaccessible to 87% of the Indian population.

Natural Language Processing (NLP) and Automatic Speech Recognition (ASR) have a long history as research topics within AI. Substantive progress made on these topics has resulted in viable systems for machine translation, spoken dialogue, sentiment analysis, and social media analysis [25]. Likely the English-speaking reader of this paper is used to receiving relevant results for search queries,

Sita saw her spouse. Edit	Sita saw her spouse. Edit
सीता ने अपने पति को देखा	சீதா தன் மனைவியை பார்த்தாள்.
seeta ne apane pati ko dekha	Cītā taṉ maṉaiviyai pārttāḷ.
(a) "Sita saw her husband."	**(b) "Sita saw her wife."**
Ram saw his spouse. Edit	Ram saw his spouse. Edit
राम ने अपने पति को देखा	ராம் தனது மனைவியை பார்த்தார்.
raam ne apane pati ko dekha	Rām taṉatu maṉaiviyai pārttār.
(c) "Ram saw his husband."	**(d) "Ram saw his wife."**

Figure 1: Results from Google Translate, accessed October 27, 2017. In (a) and (b), the sentence "Sita saw her spouse" is translated into Hindi and Tamil, respectively. In (c) and (d), the sentence "Ram saw his spouse" is translated into Hindi and Tamil, respectively. Accompanying captions show translations back into English by the <u>authors</u> (interestingly, Google translated its own Tamil translation in (b) back into the English "Sita saw his wife"). With different genders assigned to "spouse", the translations of each sentence into Hindi and Tamil are inconsistent. Most Indians will know Sita to be a common female name, and Ram a common male name. In a country with no provision for same-sex marriages, Google's translations in (b) and (c) have little chance of being correct.

recommendations of news articles to read, and even assistance from voice-controlled phone apps. Speakers of Hindi (422 million) and Tamil (60 million) face a different reality. While we are not aware of systematic studies on the quality of digital services in these languages, just a few minutes on the Internet offers a glimpse. To provide an illustration, the authors experimented with Google Translate, widely considered to be a leading service for machine translation. Figure 1 shows the translations returned for relatively simple English sentences into Hindi and Tamil. There are several "obvious" mistakes.

The success of modern NLP systems such as Google Translate owes primarily to the availability of large training corpora [5, 63]. Unfortunately, the sizes of data sets available for most Indian languages are minuscule compared to those available for major Western languages. For example, a public corpus of parallel text in 11 European languages contains tens of millions of words in each language [29]. Researchers working with Indian languages have to make do with roughly a hundredth the amount of data [31]. Thus, although NLP has been an active area of research within the Indian AI community [28], its productivity is circumscribed by the shortage of digitised data.

One strategy that has become popular in the NLP community is to use resource-rich languages as "pivots" to build applications for resource-poor languages [45]. This approach may be of independent interest to linguistics, and in the short term, can yield payoffs. However, in the long-term, we see no alternative to building systems that harvest and deliver data in Indian languages. We propose that this activity be taken up as a serious pursuit by the AI community. For example, ASR and the creative use of crowdsourcing could provide channels for digitising linguistic data. Curiously, these

areas, themselves, bemoan a lack of data and resources. Like NLP, the field of ASR has also demarcated under-resourced languages as a special topic [8]. In a recent survey, Pavlick *et al.* (2014) identify languages that are good bets for linguists to study, since they can find translators on Amazon Mechanical Turk.

Our proposal goes in the *opposite* direction, taking the bridging of India's linguistic divisions as non-negotiable, rather than the usage of convenient tools. In early stages, the emphasis need not be on complex tasks such as translation. The path to robust digital local language ecosystems could be paved by bringing more content in each language into the digital domain, providing services such as search and speech interfaces, and same-language subtitling in videos to improve functional literacy.

2.2 Structuring and Mining Public Data

Every department of the government generates records that are available in the public domain. In 2005, the Indian government passed the "Right to Information" act , which enables individuals to query governmental organisations for particular types of information. This facility—a positive step towards bringing transparency—has already been used to good effect by individuals and civil society. Yet, to make accountability and efficiency *intrinsic* to public-related offices, it is necessary to build pipelines that deliver *structured* data. In this regard, it is instructive to consider Berners-Lee's 5-star categorisation of open data [7], which is reproduced in Table 1.

In the lowest category is any data that is available on the Internet under an open licence, regardless of format (scan, picture, table) and encoding. Naturally, data is more directly usable when it is structured (for example, provided as a table, rather than natural language text) and linked with other relevant sources. Reliable, structured data is the foundation on which relevant applications and services can be built. It is clear that going *forward*, the design of data systems must aspire for a 5-star rating. Interestingly, there also emerges an excellent opportunity for AI when we look *backward*.

Even if the majority of legacy data does not meet even the 1-star criterion (being available in digital form, and accessible through the Internet), there is a substantial amount of data, especially from the last few decades, which does. This data can contain valuable information, which, unfortunately, will remain hidden unless the data can be processed into structured form. The nature and scale of

the data makes it impractical for human annotators to undertake the structuring exercise—but this is certainly something within the reach of modern AI techniques (such as computer vision and NLP). We delve into the details of a specific case study to illustrate that (1) even "macro" patterns in various data sets are often not known, and (2) gleaning them can provide invaluable inputs for course correction and policy making.

It is common knowledge that legal cases in India can be stuck in court, at various stages of appeal, for years on end [34]. We focus on cases related to income tax, which suffer judicial delays in spite of having dedicated appellate authorities: Assessing officer, Commissioner of Income Tax Appeals (CIT(A)), and Income Tax Appellate Tribunal (ITAT) [18]. Appeals from ITAT go to the High Court (HC), and thereafter the Supreme Court (SC). Table 2 shows the number of appeals and the dispute amounts locked up at different levels of litigation as of March 2015.

In general one would expect cases with *larger* dispute amounts to be appealed to higher levels, but curiously, the last column in the table shows otherwise. Observe that while the average dispute amount at ITAT is more than double that at CIT(A), it *drops* by a third when proceeding from ITAT to HC, and further from HC to SC. Clearly, understanding this trend would be key to devising measures to reduce delays at the various appellate levels. Possible explanations include (1) that the *government* is the more frequent appellant at levels beyond ITAT, for reasons of establishing precedent, and (2) there has been a large volume of cases filed over the last ten years, and these are still pending at lower levels and influencing the averages. It would *seem* a relatively straightforward matter to verify if either of these explanations is correct, but surprisingly, answers to the simple questions listed below are yet unknown!

- What fraction of cases are initiated by taxpayers and the government respectively at each level of appeal?
- In what fraction of cases are taxpayers and the government successful at each level of appeal?
- What is the average dispute amount in taxpayer appeals and government appeals?
- What is the average pendency of a case from assessment to the final resolution of the dispute?

To begin with, the ITAT and SC have independent websites that store information in different formats; until about a decade back, their sites were not even accessible through web search. Indian Kanoon (https://indiankanoon.org/), which developed specialised scrapers and offered free search services, has now become an indispensable accessory to legal research in India. Yet, even when relevant documents (such as ITAT judgements) are retrieved, their

Table 1: 5-star categorisation of open data, reproduced from the web page maintained by Berners-Lee (2010).

★	Available on the web (whatever format) but with an open licence, to be Open Data
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

Table 2: Appeals at different levels of litigation [20] (L = lakh = 10^5; C = crore = 10^7).

Appellate authority	Number of appeals	Amount in dispute (Rs)	Average per case (Rs)
CIT(A)	2.32 L	3.84 LC	1.6 C
ITAT	37,506	1.45 LC	3.9 C
HC	34,281	37,684 C	1.09 C
SC	5,661	4,654 C	82 L

lack of structure remains a hurdle. Typical ITAT judgements have multiple mentions of Rupee amounts; the only way to extract the dispute amount is from a description in natural language. For example, in *M/S Jain Furnishing vs. ACIT* (accessed November 10, 2017, https://indiankanoon.org/doc/78538052/), the dispute amount is the sum of the two amounts mentioned in the following sentence:

> "The assessee in this appeal challenged the addition of Rs. 15,609/- on account of municipal taxes and addition of Rs. 4,80,000/- disallowing part of the rent."

While it would not be trivial, it certainly appears feasible to train an NLP method to extract relevant fields such as the dispute amount from tax judgements, especially if domain knowledge can also be exploited. This simple technical intervention could eventually help identify blockages in the tax appeal hierarchy, and save precious time and resources. Similar opportunities abound in other areas of India's legacy data. For example, several opportunities in the political sphere are explored at the Trivedi Centre for Political Data (accessed November 11, 2017, https://tcpd.ashoka.edu.in/new-about-us/).

2.3 Healthcare

Access to quality health care in developing countries is a challenge that AI technologies have the potential to alleviate greatly. One of the main problems in this sector is a shortage of skilled medical personnel willing to serve away from cities [51]. The threshold set by the WHO for a country's healthcare workforce ranges from 22.8 to 59.4 skilled health workers per 10,000 population; India stands at an estimated 15.2 [23].

Modern AI facilitates ways to augment the capabilities of scarce personnel, and to some extent, offset the absence of regular lab facilities. For example, Gann et al. (2017) have demonstrated that the recurrence of prostate cancers can be predicted using features that human pathologists are typically not trained to observe. Similarly, Beck et al. (2011) apply computational methods to extract and utilise newer, more effective features for the prognosis of breast cancer. Yet another success of computational pathology is in the development of a software-controlled microscope that can detect malaria at expert-level accuracy in the field [19]. Neonatal sepsis is a large contributor to neonatal mortality [53]. Studies show that time-series data from standard non-invasive measurements, such as heart rate and respiration over the first few hours of the life of a preterm baby, can predict morbidity with accuracy comparable to invasive (and often expensive and unavailable) lab tests [54].

Data-driven algorithms can also inform epidemiological analysis to understand disease burden and response. The POSEIDON study [52] was a well-conceived exercise that recorded data from clinics in 880 cities and towns in India on a *single* day. Even just a preliminary analysis of this data, gathered from over 200,000 patients, reveals patterns in the frequency of visits to health facilities across gender and age groups, categories of illnesses, etc. There are also clear differences from similar data sets gathered in other countries such as Sri Lanka and Singapore. It follows that large-scale data analysis can provide non-trivial inputs to healthcare policy. AI can contribute the technology to digitise health records using automated capture methods such as IoT-enabled medical devices and app-based forms with location and image-based inputs. The objective would be to construct pipelines that deliver authentic and accurate data, with minimal human intervention.

the pilferage of medicines and theft of equipment [14]. Solutions for tracking inventory and personnel, based on computer vision, could possibly cut down on such losses in delivery.

Although we have singled out healthcare as an illustrative "vertical" in this section, it must be noted that both problems and effective solutions tend to spill over boundaries. For example, the root causes of poor health in a population could include limited access to information and education, poor quality of service delivery due to lack of infrastructure and corruption, and a debt trap from high out-of-pocket expenses. As a general strategy, it would be advisable to understand the dependencies between various problems before rushing into solutions.

3 RISKS OF AN AI-CENTRIC APPROACH

That AI can contribute to development in numerous (often unconventional) ways creates a climate of hope and optimism. However, it would be naïve not to anticipate and forestall the potential risks of AI-driven growth. In this section, we raise the main concerns that emerge from India's socio-economic context.

Displacing workers. India is no exception to the global AI wave, which is beginning to uproot workers from their jobs [11]. A recent study by McKinsey and Company (2014) estimates that 6–8 million workers "currently employed in routine clerical, customer service, and sales jobs could be affected by advancements in machine learning and natural language interfaces (speech recognition)." A loss of jobs at this scale can have an impact on economic well-being for a large number of people who may be dependent on these wage-earners, an important consequence for a middle-income country trying to raise a large number of citizens out of poverty. India's acclaimed IT industry is already feeling the pinch of automation [58], suggesting that a crisis triggered by job losses could hit the population over the next few years. Other side effects of AI might take longer to manifest.

Reinforcing social discrimination. The caste system in India is a social hierarchy with historical roots. Sadly it continues to perpetuate discrimination in subversive and invisible ways, affecting wages [4], employment [36], imprisonment rates [46], and access to credit from banks [30]. With the advent of AI, it has become a growing concern that data-driven algorithms can pick up biases from the data they are fed. For example, in the United States, algorithms for assessing recidivism rates [2] are suspected to show racial biases [16]. Markers of caste and religion are present in names and addresses, and can easily affect data-driven algorithms that might be used to assess applications for jobs, loans, or bail [33]. An experiment conducted some years back by Banerjee *et al.* (2009) found evidence of caste-based discrimination in call-centre job applications. Even if we presume that the decisions, in this case, were made by human evaluators, it is a cause for concern if these decisions are eventually used to train an algorithm for screening applications.

Amplifying gender inequality. The number of Internet users and the number of mobile Internet users in India are both expected

to grow—to 420 million and 300 million, respectively, in 2017 [26]. Mobile phones are the primary access point to the Internet, particularly in rural India, where 60% of Internet access is through mobile phones. While the penetration of mobile phones seems overall a boon for AI, it could unwittingly amplify the gender disadvantage. Women in South Asia are 38% less likely to own a mobile phone than men; when overlaid with patriarchal and misogynistic social norms, this means the real access rate could be even lesser [24]. Consequently the reach of AI may become segmented along gender lines (as also other divisions arising from economic and geographical barriers).

A second worry is that gender ratios in India's software industry are heavily skewed at all levels [32]. Hence, there is a real risk that the AI to be consumed by the entire population will be produced with a strong male bias. This imbalance could create undesirable long-term consequences [59].

Excluding the disadvantaged through targeting. The high costs of developing AI-based applications may mean that the initial impetus will come from private corporations. It is natural for corporations to seek revenues from areas in which profit pools are large, with no particular obligation to address socially-relevant issues such as equitable access. Consequently, the needs of the less-profitable may not be considered. The example from Figure 1 is instructive: it is unlikely that Google will prioritise its Tamil→Hindi translation engine as high as its English→Mandarin engine. When commercial interests are overlaid with AI-based marketplaces, there is a risk that the poor are further marginalised. A recent essay by Calo and Rosenblat (2017) serves as a compelling account of this worrying possibility.

4 AI FOR DEVELOPMENT: STEPS AND SAFEGUARDS

In this section, we propose some guiding principles for the construction of a robust AI ecosystem in India.

Neither automobile engines nor air conditioners could have been built without the humble thermometer. At this juncture, it is imperative to build the instruments to *measure* India's "vital statistics", in order that they can thereafter be *improved*. To be effective, AI needs access to relevant data in the digital domain. As already outlined in Section 2.2, we recommend that the construction of 5-star data pipelines be taken up on a priority basis. The government's "Digital India" (accessed October 6, 2018, https://www.digitizeindia.gov.in/) and "Open Government Data" initiatives (accessed October 6, 2018, https://data.gov.in/) are welcome steps in this direction. In addition to public data from governmental departments, it would also be useful to create locally-relevant public open sets pertaining to language, health, crops, marketplaces, and so on. In some cases, AI technologies such as computer vision and crowdsourcing could themselves be deployed to seed the effort of creating such data sets.

It would neither be effective nor sustainable if the activity of developing AI-based solutions is confined to a small number of people and places. It is essential that a broader section of the population—especially women, linguistic minorities, and rural communities—be actively trained to create and maintain AI systems for their own needs. Our example of Google's incorrect translations in Figure 1

remains illustrative. Clearly such mistakes can be rectified more quickly and effectively by local speakers who are aware of existing gender biases in their own languages [10]. However, they will need both the data and the technical knowledge to develop and maintain their own translation engines. The proposal put forth by Jain (2002)—to actively complement the Nehruvian. top-down model of knowledge-generation and dissemination, with the Gandhian, bottom-up model—is of especial relevance to the growth of the AI knowledge network. The open source movement has been reasonably successful in India, and can be expanded for the development of AI libraries, standards, and APIs.

India enjoys the advantages of having an established university system and a well-trained workforce. However, the supply of knowledge and skill are no match for the demand created by a large, diverse, and developing country. Flagship demonstrations such as Deepmind's AlphaGo program [56], especially if they can be situated in the Indian context, can excite young minds to pursue careers in AI. So also would the publication of interesting data sets and the organisation of competitions. Domestic centres of excellence in research could provide leadership not just in core AI technologies, but also in interdisciplinary areas. If AI is the new electricity, society would need not only electrical engineers, but also electricians. Measures to train a large workforce to build applications using vision, speech, and so on would be a positive step, which may also help by absorbing some of the shock created by job losses.

Industry, especially startups, will play a vital role in identifying and realising the benefits of AI across diverse sectors. India has a thriving tech entrepreneurship ecosystem, with access to talent, capital, and large markets. There are about 300 startups in India with a focus on AI, as of May, 2017 [44], with over USD 100 million invested in them since 2014 [55]. This number, however, is low in comparison to countries like the US and China, where investments total over USD 4 and 3 billion, respectively. Lack of data sets and talent are both challenges that startups will have to negotiate; closer collaboration with universities could help in the latter respect. Startups that are constrained to keep risk low can focus on high-volume, low-margin sectors. For example, even a 5% reduction in raw-material wastage, power consumption, or rejection rate, can be substantial in the manufacturing industry.

In step with the growth of AI, India will also have to evolve regulatory mechanisms such as safety and quality standards; legal frameworks addressing data security, privacy, and liability; and ethics review committees.

5 CONCLUSION

This paper takes a first step towards defining the contours of AI in India. Our foremost observation is that AI brings with it both *opportunities*, which tend to be apparent and enticing, and *risks*, whose effects might take longer timescales to manifest themselves. With thoughtful planning and management, we believe that not only can AI have a net positive effect on India's development, indeed it can help leapfrog traditional hurdles to growth.

With such high stakes involved, it is imperative that the emergence of AI in India be subjected to rigorous *academic* study. We primarily intend this paper to be a seed for such work. We hope that the template provided, as also the numerous references, can

help the interested researcher get started on more detailed investigations. Admittedly our coverage of areas (we have only surveyed healthcare, and that too briefly) is insufficient. For a more comprehensive listing of verticals, we refer the reader to other recent reports [40, 47, 57], which devote full sections to agriculture, transportation, education, urban planning, security, employment, entertainment, manufacturing, robotic automation, and environment. The risks we have highlighted in Section 3 also merit more specific and quantitative analysis.

ACKNOWLEDGEMENTS

REFERENCES

[1] Nisha Agrawal. 2016. Inequality in India: what's the real story? *World Economic Forum* (2016). Accessed October 26, 2017, www.weforum.org/agenda/2016/10/inequality-in-india-oxfam-explainer.
[2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2017. Machine Bias. *Pro Publica* (2017). Accessed October 10, 2017, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[3] Abhijit Banerjee, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. Labor market discrimination in Delhi: Evidence from a field experiment. *J. Comparative Economics* 37, 1 (2009), 14–27.
[4] Biswajeet Banerjee and J. B. Knight. 1985. Caste Discrimination in the Indian Urban Labour Market. *Journal of Development Economics* 17, 3 (1985), 277–307.
[5] Michele Banko and Eric Brill. 2001. Scaling to Very Very Large Corpora for Natural Language Disambiguation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 26–33.
[6] Andrew H Beck, Ankur R Sangoi, Samuel Leung, Robert J Marinelli, Torsten O Nielsen, Marc J van de Vijver, Robert B West, Matt van de Rijn, and Daphne Koller. 2011. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine* 3, 108 (2011), 108ra113.
[7] Tim Berners-Lee. 2010. Linked Data. (2010). Accessed October 28, 2017, www.w3.org/DesignIssues/LinkedData.html.
[8] Laurent Besacier, Etienne Barnard, Alexey Karpov, and Tanja Schultz. 2014. Automatic Speech Recognition for Under-resourced Languages: A Survey. *Speech Communication* 56 (2014), 85–100.
[9] Nicola J. Bidwell. 2016. Moving the centre to design social media in rural Africa. *AI & Society* 31, 1 (February 2016), 51–77.
[10] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Advances in Neural Information Processing Systems 29*. Curran Associates, 4349–4357.
[11] Erik Brynjolfsson and Andrew McAfee. 2016. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
[12] Ryan Calo. 2016. Robots as Legal Metaphors. *Harvard Journal of Law and Technology* 30, 1 (2016), 209–237.
[13] Ryan Calo and Alex Rosenblat. 2017. The Taking Economy: Uber, Information, and Power. *Columbia Law Review* 117 (2017), 1623–1690.
[14] Subrata Chattopadhyay. 2015. Corruption in healthcare and medicine. *Indian Journal of Medical Ethics* 10, 3 (2015), 153–189.
[15] Mike Cooley. 1995. The myth of the moral neutrality of technology. *AI & Society* 9, 1 (March 1995), 10–17.
[16] Kate Crawford. 2016. Artificial Intelligence's White Guy Problem. *The New York Times* (2016). Accessed October 4, 2017, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.
[17] Credit Suisse. 2014. Global Wealth Databook 2014. (2014). Accessed October 27, 2017, http://publications.credit-suisse.com/index.cfm/publikationen-shop/research-institute/global-wealth-databook-2014/.
[18] Pratik Datta, B. S. Surya Prakash, and Renuka Sane. 2017. Understanding Judicial Delay at the Income Tax Appellate Tribunal in India. *National Institute for Public Finance and Policy* (2017). Accessed October 28, 2017, nipfp.org.in//media/medialibrary/2017/10/WP_2017_208.pdf.
[19] Charles B Delahunt, Courosh Mehanian, Liming Hu, Shawn K McGuire, Cary R Champlin, Matthew P Horning, Benjamin K Wilson, and Clay M Thompon. 2015. Automated microscopy and machine learning for expert-level malaria field diagnosis. In *Global Humanitarian Technology Conference (GHTC), 2015 IEEE*. IEEE, 393–399.
[20] Department of Revenue. 2016. Report of the Comptroller and Auditor General of India. (2016). Accessed November 13, 2017, www.cag.gov.in/sites/default/files/audit_report_files/Union_Direct_Tax_Compliance_Revenue_Report_3_2016_Department_Revenue.pdf.
[21] Ariel Felner. 2016. The Israeli AI Community. *AI Magazine* 37, 3 (2016), 118–122.
[22] Peter H Gann, Lingdao Sha, Virgilia Macias, Neeraj Kumar, and Amit Sethi. 2017. Computer Vision Detects Subtle Histological Effects of Dutasteride on Benign Prostate Tissue. *The FASEB Journal* 31, 1 Supplement (2017), lb520.
[23] Global Health Workforce Alliance and World Health Organization. 2013. A Universal Truth: No Health Without a Workforce. *WHO Report* (2013). Accessed November 8, 2017, www.who.int/workforcealliance/knowledge/resources/ hrhreport2013/en/.
[24] GSMA. 2015. Bridging the gender gap: Mobile access and usage in low- and middle-income countries. (2015). Accessed October 10, 2017, www.gsma.com/mobilefordevelopment/wp-content/uploads/2016/02/GSM0001_03232015_GSMAReport_ NEWGRAYS-Web.pdf.
[25] Julia Hirschberg and Christopher D. Manning. 2015. Advances in natural language processing. *Science* 349, 6245 (2015), 261–266.
[26] IAMAI and Kantar IMRB. 2017. Internet in India – 2016. (2017). Accessed October 26, 2017, bestmediainfo.com/wp-content/uploads/2017/03/Internet-in-India-2016.pdf.
[27] Ashok Jain. 2002. Networks of Science and Technology in India: The Elite and the Subaltern Streams. *AI & Society* 16, 1–2 (May 2002), 4–20.
[28] Deepak Khemani. 2012. A Perspective on AI Research in India. *AI Magazine* 33, 1 (2012), 96–98.
[29] Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. (2005). Accessed November 8, 2017, https://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf.
[30] Sunil Mitra Kumar. 2013. Does Access to Formal Agricultural Credit Depend on Caste? *World Development* 43, C (2013), 315–328.
[31] Anoop Kunchukuttan, Abhijit Mishra, Rajen Chatterjee, Ritesh M. Shah, and Pushpak Bhattacharyya. 2014. Śata-Anuvādak: Tackling Multiway Translation of Indian Languages. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. European Language Resources Association, 1781–1787.
[32] Jadine Lannon. 2013. Research Initiative: Women in India's IT Industry. *The Centre For Internet & Society* (February 2013). Accessed November 13, 2017, cis-india.org/internet-governance/blog/women-in-indias-it-industry.
[33] Issie Lapowsky. 2017. One State's Bail Reform Exposes the Promises and Pitfalls of Tech-Driven Justice. *Wired* (2017). Accessed October 10, 2017, www.wired.com/story/bail-reform-tech-justice.
[34] Law Commission of India. 2014. Report No. 245: Arrears and Backlog: Creating Additional Judicial (wo)manpower. (2014). Accessed October 27, 2017, lawcommissionofindia.nic.in/reports/Report_No.245.pdf.
[35] Stephen E. Little. 1999. Science, technology and society in East Asia: Frameworks for the challenges of the next century. *AI & Society* 13, 3 (September 1999), 247–262.
[36] S Madheswaran and Paul Attewell. 2007. Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey. *Economic and Political Weekly* 42, 41 (2007), 4146–4153.
[37] Miwao Matsumoto. 1999. The .Japan problem. in science and technology and basic research as a culture. *AI & Society* 13, 1–2 (March 1999), 4–21.
[38] McKinsey and Company. 2014. India's technology opportunity: Transforming work, empowering people. (2014). Accessed October 10, 2017, www.mckinsey.com//media/McKinsey/Industries/High Tech/Our Insights/Indias tech opportunity Transforming work empowering people/MGI India tech_Full report_ December 2014.ashx.
[39] Anthony R. Measham and Meera Chatterjee. 1999. *Wasting Away: The Crisis of Mulnutrition in India*. IBRD/World Bank.
[40] Ministry of Commerce and Industry, Government of India. 2017. Artificial Intelligence Task Force. (2017). Accessed November 5, 2017, www.aitf.org.in/.
[41] Ministry of Finance, Government of India. 2017. Economic Survey 2016–17. (2017). Accessed November 5, 2017, indiabudget.nic.in/es2016-17/echapter.pdf.
[42] Ministry of Home Affairs, Government of India. 2011. Census of India 2001: Data on Language. (2011). Accessed November 5, 2017, www.censusindia.gov.in/Census_Data_2001/Census_Data_ Online/Language/data_on_language.aspx.
[43] Ministry of Human Resource Development, Government of India. 2014. Education for All: Towards Quality with Equity. (2014). Accessed November 5,

2017, mhrd.gov.in/sites/upload_files/mhrd/files/upload_document/EFA-Review-Report-final.pdf.

[44] Mint. 2017. 10 standout start-ups taking an AI leap in India. (May 2017). Accessed November 8, 2017, www.livemint.com/Leisure/u7M3e5ymwmGf6QRLaXBoAJ/10-standout-startups-taking-an-AI-leap-in-India.html.

[45] Preslav Nakov and Hwee Tou Ng. 2012. Improving Statistical Machine Translation for a Resource-Poor Language Using Related Resource-Rich Languages. *Journal of Artificial Intelligence Research* 44 (2012), 179–222.

[46] National Crime Record Bureau. 2016. Prison Statistics India 2015. (2016). Accessed October 10, 2017, http://ncrb.gov.in/statpublications/psi/Prison2015/Full/PSI-2015- 18-11-2016.pdf.

[47] NITI Aayog, Government of India. 2018. National Strategy for Artificial Intelligence. (2018). Accessed October 6, 2018, http://niti.gov.in/writereaddata/files/document_publication/ NationalStrategy-for-AI-Discussion-Paper.pdf.

[48] Ramya Parthasarathy and Vijayendra Rao. 2017. Deliberative Democracy in India. *World Bank Policy Research Working Paper* (2017). Accessed October 27, 2017, documents.worldbank.org/curated/ en/428681488809552560/pdf/WPS7995.pdf.

[49] Ellie Pavlick, Matt Post, Ann Irvine, Dmitry Kachaev, and Chris Callison-Burch. 2014. The Language Demographics of Amazon Mechanical Turk. *Transactions of the Association for Computational Linguistics* 2 (2014), 79–92.

[50] Áron Perényi. 2014. Are theories applicable across different contexts? A cross-national comparative analysis through the lens of firm life-cycle theory in the ICT sector. *AI & Society* 29, 3 (August 2014), 289–309.

[51] Bhaskar Purohit and Tathagata Bandyopadhyay. 2014. Beyond job security and money: driving factors of motivation for government doctors in India. *Human Resources for Health* 12, 1 (2014), 12.

[52] Sundeep Salvi, Komalkirti Apte, Sapna Madas, Monica Barne, Sushmeeta Chhowala, Tavpritesh Sethi, Kunal Aggarwal, Anurag Agrawal, and Jaideep Gogtay. 2015. Symptoms and medical conditions in 204,912 patients visiting primary health-care practitioners in India: a 1-day point prevalence study (the POSEIDON study). *The Lancet Global Health* 3, 12 (2015), e776 – e784.

[53] MJ Sankar, SB Neogi, J Sharma, M Chauhan, R Srivastava, PK Prabhakar, A Khera, R Kumar, S Zodpey, and VK Paul. 2016. State of newborn health in India. *Journal of Perinatology* 36 (2016), S3–S8.

[54] Suchi Saria, Anand K Rajani, Jeffrey Gould, Daphne Koller, and Anna A Penn. 2010. Integration of early physiological responses predicts later illness severity in preterm infants. *Science translational medicine* 2, 48 (2010), 48ra65.

[55] Sriram Sharma. 2017. Here's why India is likely to lose the AI race. (Aug 2017). Accessed November 8, 2017, factordaily.com/artificial-intelligence-india/.

[56] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

[57] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. 2016. *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel.* Technical Report. Stanford University.

[58] Samanth Subramanian. 2017. India Warily Eyes AI. *MIT Technology Review* (October 2017). Accessed November 13, 2017, www.technologyreview.com/s/609118/india-warily-eyes-ai/.

[59] Joan Truckenbrod. 1993. Women and the social construction of the computing culture: Evolving new forms of computing. *AI & Society* 7, 4 (December 1993), 345–357.

[60] Pradeep Varakantham, Bo An, Bryan Low, and Jie Zhang. 2017. Artificial Intelligence Research in Singapore: Assisting the Development of a Smart Nation. *AI Magazine* 38, 3 (2017), 102–105.

[61] Shashi Shekhar Vempati. 2016. India and the Artificial Intelligence Revolution. (2016). Accessed November 5, 2017, carnegieindia.org/2016/08/11/india-and-artificial-intelligence-revolution-pub-64299.

[62] Jonathan Woetzel, Anu Madgavkar, and Shishir Gupta. 2016. A new emphasis on gainful employment in India. *Report, McKinsey Global Institute* (2016). Accessed October 27, 2017, https://www.mckinsey.com/featured-insights/employment-and-growth/a-new-emphasis-on-gainful-employment-in-india.

[63] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. (2016). Accessed October 29, 2017, arxiv.org/abs/1609.08144.