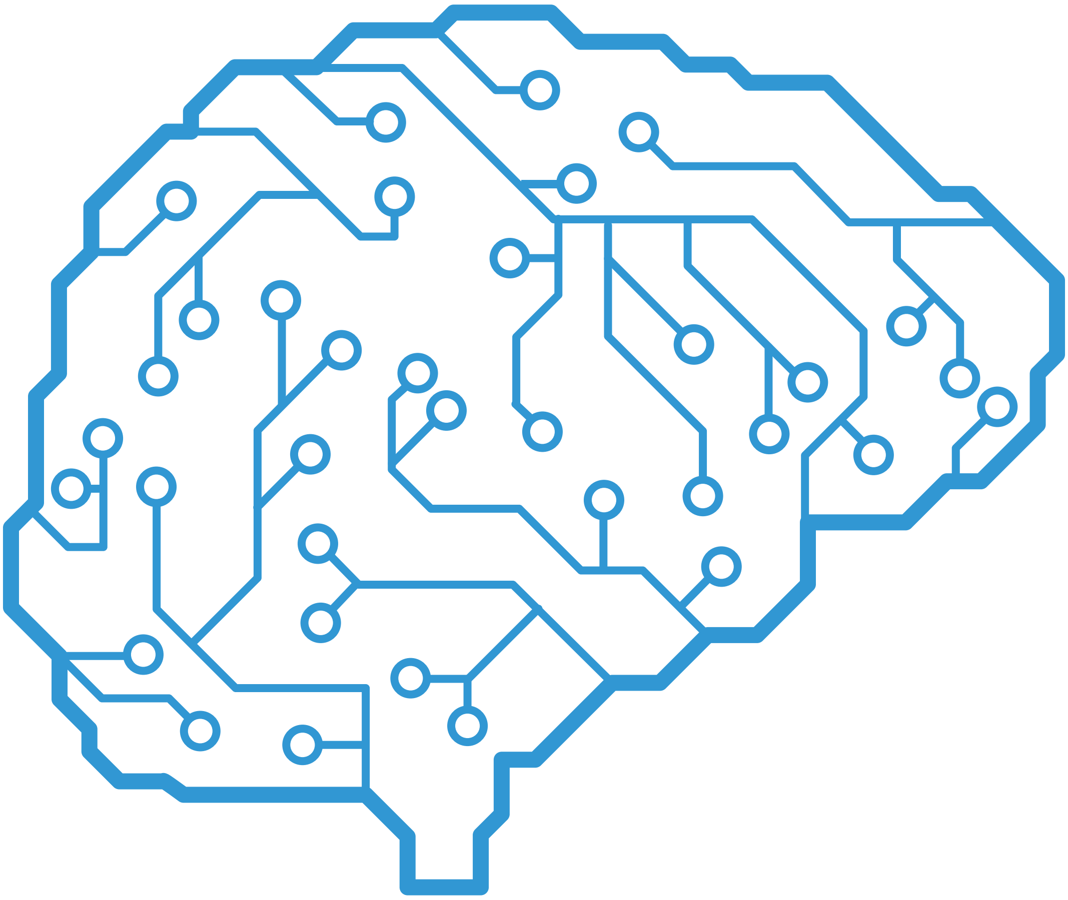
December 2018

INSTITUTE FOR TRANSFORMATIVE TECHNOLOGIES



ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS FOR HUMAN DEVELOPMENT

Separating facts from hype on where AI and data can genuinely help, and where it is a distraction

Shashi Buluswar, Roger Sathre, Pranav Sridhar and Rahul Aluri

Authors

Shashi Buluswar, Roger Sathre, Pranav Sridhar and Rahul Aluri

Attribution

Please cite this work as follows: "Institute for Transformative Technologies (ITT), December 2018. "Artificial Intelligence and data analytics or human development: Separating facts from hype on where AI and data can genuinely help, and where it is a distraction"

For more information

info@transformativetechnologies.org | www.transformativetechnologies.org

Table of Contents

Executive Summary

In recent years, the rapidly growing presence of Artificial Intelligence (AI) in virtually every aspect of life in the industrialized world has led to important questions on how AI can help achieve the 2030 Sustainable Development Goals (SDGs). The conversations on AI in the context of human development are nascent, with many interesting ideas being proposed. However, the discussions to date have largely been abstract, with little specificity about where big data and AI can help. Indeed, we believe there is a significant knowledge gap between the AI community and the development community, which needs to be closed in order to separate fact from hype, and to ensure resources are dedicated to initiatives with a meaningful likelihood of impact.

To that end, this article identifies the most important interventions required to achieve the SDGs, and assesses their dependence on data analytics and AI. Our analysis finds that robust data analytics can play a meaningful role in achieving the SDGs. However,

- About half of the major interventions required to achieve the SDGs do not need sophisticated data analytics; rather they depend on infrastructure (for energy access), physical inputs (for increased crop production), innovative business models (for education and healthcare), and other on-the-ground programs. While data can help inform and optimize these programs, it constitutes only a small portion of the overall solution space.
- There are a number of important interventions which depend heavily on data and analytics, but the vast majority of them can be implemented with conventional analytics solutions, the likes of which have been in commercial use for at least two decades in industrialized countries.
- Advanced AI can offer powerful solutions in a small number of specific, targeted areas such as automated health diagnostics. In most other cases, AI risks being an overkill and a distraction.

About half of the major interventions required to achieve the SDGs do not need sophisticated data analytics; rather they depend on infrastructure, physical tools, innovative business models, and other on-the-ground programs.

To fully implement the interventions that do rely on data, as well as for long-term development, transparency and accountability, countries need to invest in a robust data infrastructure. To measure the breadth and depth of a country's digital environment, we introduce the Data Density Index (DDI), a composite metric that assesses the volume and variety of data generated across platforms built by both governments (for key public services) and private companies (e.g., smartphones and apps).

Based on the available evidence, the DDI for most developing countries ranks them as "data deficient". This suggests that despite the telecom revolution, much needs to be done before decisions and interventions are informed by robust, granular data.

In that context, India's Aadhar and India Stack initiatives represent a powerful model of digital and data inclusion. If such systems are implemented in other parts of the developing world, they can make the hype of data analytics—and perhaps even AI—become reality, over time. In many ways, the timing may be ideal for emerging economies to make robust investments in their data infrastructure.

As such, institutions aiming to impact the SDGs have four options: (i) focusing on more direct interventions (e.g., irrigation and seed hybridization for agricultural development); (ii) investing in conventional analytics solutions to improve decision-making on direct interventions; (iii) building data infrastructures for the long haul; and (iv) betting on new-generation AI solutions. Depending on the context, each of the first three options can be valuable and can be implemented in tandem; however, we believe the AI option is the least likely to lead to impact in the timeline for the SDGs.

1. Introduction

The current conversation on AI for human development

Over the past century, Artificial Intelligence has possibly been the single most misconstrued, overhyped, and even feared technological construct.

Perhaps it is because AI makes for ever-reliable science fiction plotlines; perhaps because the concept evokes, at once, humanity's limitlessness and its ultimate fallibility; or maybe it is because we just don't understand it very well.

In 2002, one of this report's authors made the following observation: "Artificial Intelligence has come in and out of vogue more times than Madonna in the past 20 years: it has been hyped and then, having failed to live up to the hype, been discredited until being revived again.[1]" Sixteen years hence, while Madonna appears to be faithfully following that script, AI seems to have established a permanent foothold, at least in the industrialized world.

While the number of applications of AI—along with a host of legitimate fears—is growing at a dramatic pace in the industrialized world, the conversation about how it can address problems related to poverty in low-income countries is relatively nascent. In the context of the developing world, the cycle of hype, hope and disillusionment appears to have only just begun.

While a number of potential data-driven solutions exist at concept and early stages, there has been very little in the way of deployment at any meaningful scale.

Today, AI is used to hyper-customize a range of products and services to our exacting preferences; to process language and speech, and assist us with a broad range of tasks such as voice-activated instructions to our smartphones, mapping and directions, and telling us what the weather will be like; in smart robotic manufacturing lines; and in self-driving cars which seemingly interpret and navigate their environments better than human drivers can. Where this all will lead in the years and decades to come is, of course, a topic of much animated debate.

[1] "The Return of Artificial Intelligence: Is AI finally ready for business?" C. Booth and S. Buluswar, McKinsey Quarterly, 2002.

We strongly believe that the current conversations on the human development value of AI are highly abstract, have limited relevance to actual on-the-ground problems, and are vulnerable to far too many "silver bullet" projections about how AI will help address humanity's most pressing problems and achieve the Sustainable Development Goals (SDGs) by the 2030 deadline.

An influential report from a 2017 "AI for Good" summit[2] asserted "a firm belief that AI will help to solve some of the most pressing challenges to our planet and its people," and stated that "AI will be central to the achievement of the Sustainable Development Goals and could help solve humanity's grand challenges."

A 2018 report[3] claimed that "there is enormous potential to create AI-enabled 'game changers' in which the application of AI, often in combination with other Fourth Industrial Revolution technologies, has the potential to deliver transformative solutions."

The list on Page 6, while hardly exhaustive, is a fair representation of the breadth and types of AI-based solutions being proposed for human development and the SDGs. As the examples show, AI researchers are taking on a broad range of challenges for the benefit of disadvantaged and underserved communities. However, our examination of the AI solutions being discussed at the major forums found that only a small minority of them are positioned for meaningful impact.

According to our assessment, most such proposed AI solutions face at least one of the following challenges: first, the term AI itself is incorrectly applied to refer to anything related to data-driven analytics; second, they appear to assume that simply improving quality of information will lead to improved execution and greater impact; and third, many of them are better suited to supplement genuine high-priority interventions, rather than be high-priority interventions unto themselves.

Indeed, our (admittedly harsh) assessment is that the vast majority of the solutions currently being discussed are neither necessary nor sufficient for meaningful impact; and investing in them without a deeper understanding of where data/AI can genuinely help risks being a distraction from critical solutions that could significantly improve SDG outcomes. The key problem is the severe knowledge divide that lies between experts in AI and experts in the various human development domains, with each side needing to learn much more about the other domain before effective solutions can be jointly developed.

---

[2] "AI for Good Global Summit Report", Geneva Switzerland, 2017.
[3] "Fourth Industrial Revolution for the Earth: Harnessing Artificial Intelligence for the Earth" PwC, 2018.

To be sure, data analytics can be used in a variety of ways to address the most challenging human development problems. At the same time, foundational digitization platforms like India's Aadhar and India Stack initiatives are beginning to demonstrate the long-term value in investing in public data infrastructures, so that developing economies can leap-frog towards a data-driven digital economy.

However, identifying the best opportunities for using data analytics and AI requires carefully analyzing the underlying problems, prioritizing the various interventions required to address those problems, and then understanding the data dependencies of those interventions. To that end, this article explains the essentials of AI and data-driven analytics, analyzes the solution space of a number of human development problems, and identifies the most valuable contributions that AI and data analytics can make.

Examples cited of various potential uses of AI for human development include:

Health
medical diagnostic devices with automated decision-making; models to predict cholera outbreaks and distribute fresh water and vaccines; tracking of clinicians' movements to ensure compliance with hygiene protocols; arm bands to monitor the nutritional status of children; analysis of medical imaging to identify tumors and other anomalies; prediction of outbreaks of infections; genetic analysis to determine vulnerability to diseases

Food security
satellite imagery to monitor and predict agricultural yield patterns; hyper-local weather analysis to inform farmers of optimal timing; precision-control agricultural systems to optimize crop production; predicting food prices from Twitter; microdrones for crop pollination

Energy Access
drones for inspecting power lines; predicting solar flares to protect power grids; optimized energy system forecasting and management

Humanitarian assistance
modeling and predicting extreme weather events; using satellite imagery to monitor flood damage and displaced populations

Communication
real-time translation between multiple languages

Education
learning system with personalized content and pace; open-access platform for free, worldwide dissemination of scientific advances to scientists, researchers, funders and policy-makers

Peace and security
monitoring seismic activity to ensure compliance with the nuclear test treaty; mapping discrimination against refugees in Europe; recognizing rescue attempts from shipping data in the Mediterranean

Conservation
utilizing data on fishing vessels to identify violations of ocean conservation agreements; detecting fires in Indonesian rainforests; detecting and monitoring harmful algal blooms; predicting bird habitat and migration patterns

Other topics
self-driving cars; robots to guide people around buildings; automated analysis of demographics, voting patterns, etc. from Google Streetview® images

2. Understanding AI

Definitions, evolution, and the distinction between AI and conventional analytics

Over the decades, one constant about AI has been its abstract and aspirational definition: a computational mechanism that mimics human sensory, cognitive, interactive and manipulative abilities, to the point of being indistinguishable from[4] (or even surpassing) human capabilities.

During that period, however, the specific computational tools described as AI in industry parlance have evolved dramatically.

Through this evolution of AI and its definitions across computational generations, what was considered AI in one generation is inevitably considered no more than "conventional" analytics in the next generation.

The mystery and awe inspired by the hidden mechanics of the most powerful "intelligent" tools of the day invariably lead to the eventual recognition that they are nothing more than robust mathematical models which have maximized the potential of available data and computing power.

The advances in AI over the decades has been driven by three factors.

- First, computational power and memory, following Moore's Law, have become exponentially more compact and inexpensive, making it possible to process significantly more data and calculations in a tiny fraction of the time and space it took just a couple of decades back.

- **1960s-70s:**
  Relatively simple expert systems (essentially if-then-else rules for decision-making)
- **1980s-90s:**
  First-generation neural networks (sophisticated statistical tools for building complex models and learn patterns from large amounts of data)
- **Today:**
  Deep learning algorithms (highly sophisticated variations of tools like neural networks, using massive amounts of data to model very complex patterns).

[4] Among the first canonical definitions of artificial/computational intelligence was the Turing Test (proposed by the mathematician Alan Turing in his seminal 1950 paper "Computing Machinery and Intelligence", in which a computing system passes the test if a human evaluator cannot distinguish its responses from that of a human, in a natural language conversation.

- Second, the sheer volume of transactional data captured about individuals and their interactions with other individuals, vendors, institutions and their environment has increased by several orders of magnitude.

- Third, the proliferation of an extraordinary range of sensors capturing digitized data—from GPS trackers, to cameras, IoT devices, satellites, drones and medical diagnostic devices—has enabled the analytical engines to incorporate hitherto unfathomable data types, and model extraordinarily complex patterns and relationships across a large (and ever-growing) number of dimensions.

Going back a few decades, the financial services sector was an early adopter of AI and data analytics to improve risk profiling, product pricing and customer targeting. Exhibit 1 illustrates the methods used by banks, credit card companies and insurance providers to move from relatively simple statistical models to early-generation AI in the 1990's and early 2000s.

This shift was precipitated by the increasing availability of numeric/digital data and the emergence of the data aggregation industry. The principal analytical tools—such as neural networks—had hitherto been only used in AI research settings.

Within a few years, the vast majority of financial services providers in the industrialized world adopted such tools, which quickly became accepted as conventional analytics, indeed even the bare minimum level of sophistication required to function in the industry.

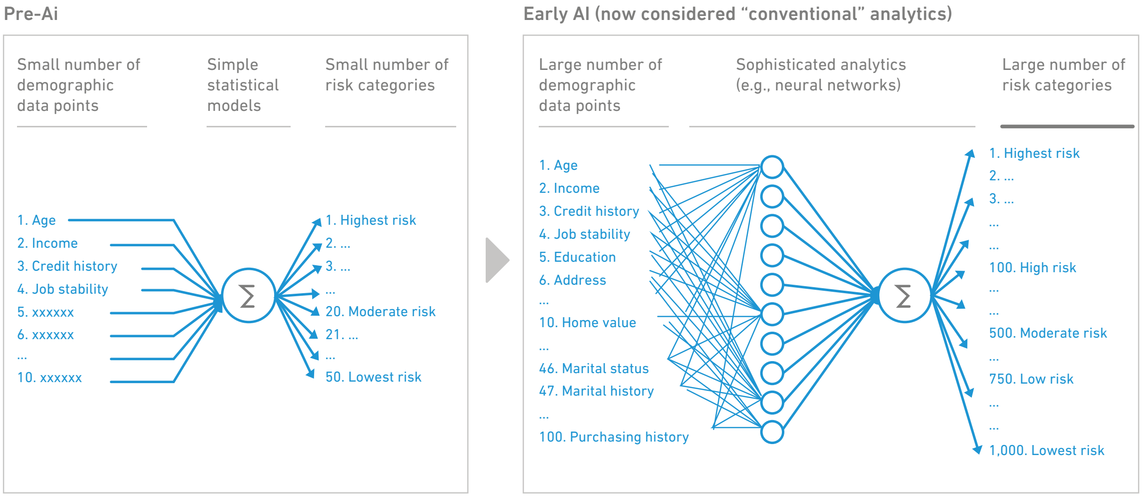


Exhibit 1. The financial services industry was among the early adopters of sophisticated data analytics and early-generation AI for risk scoring. Before the availability of a breadth of data types and data points, financial services companies used rudimentary models to determine customer risk. As digitized numerical data became available, early movers in the industry quickly adopted neural networks and other analytical/AI tools hitherto used only in academic research, to make much more precise predictions about customer profiles. By the turn of the century, such tools became commonplace in a number of industries, and were no longer considered "AI".

The period between 2005 and 2010 marked the advent of the "big data" era[5,6,7]. With the dramatic increase in data sources enabled by smart phones, other smart devices (e.g., fitness trackers), social networking platforms, increased sophistication of search engines like Google, and continued growth of the data aggregation industry, the amount of data—as well as the variety of data—available about consumers increased by orders of magnitude.

Not surprisingly, this allowed financial services providers (and, increasingly, companies in many other sectors) to become much more sophisticated in identifying potential customers, understanding their profiles and tendencies, and marketing to them with highly customized advertisements, pricing and post-sales customer relationship management.

[5] "A Very Short History Of Big Data", G. Press, Forbes, May 2013.
[6] "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society", R. Bryant, R. Katz, E. Lazowska, December 2008.
[7] Note that the phrase "big data" was coined about a decade earlier, reportedly in "Application-controlled demand paging for out-of-core visualization" by M. Cox and D. Ellsworth (Proceedings of the IEEE 8th conference on Visualization, October 1997).

In order to fully utilize the newfound breadth and volume of data, the analytical engines also needed to become more powerful, which led conventional neural networks to evolve into "deep learning" neural networks capable of extracting many more patterns and relationships. Exhibit 2 illustrates the power of the new-generation AI tools over those from the previous generation.

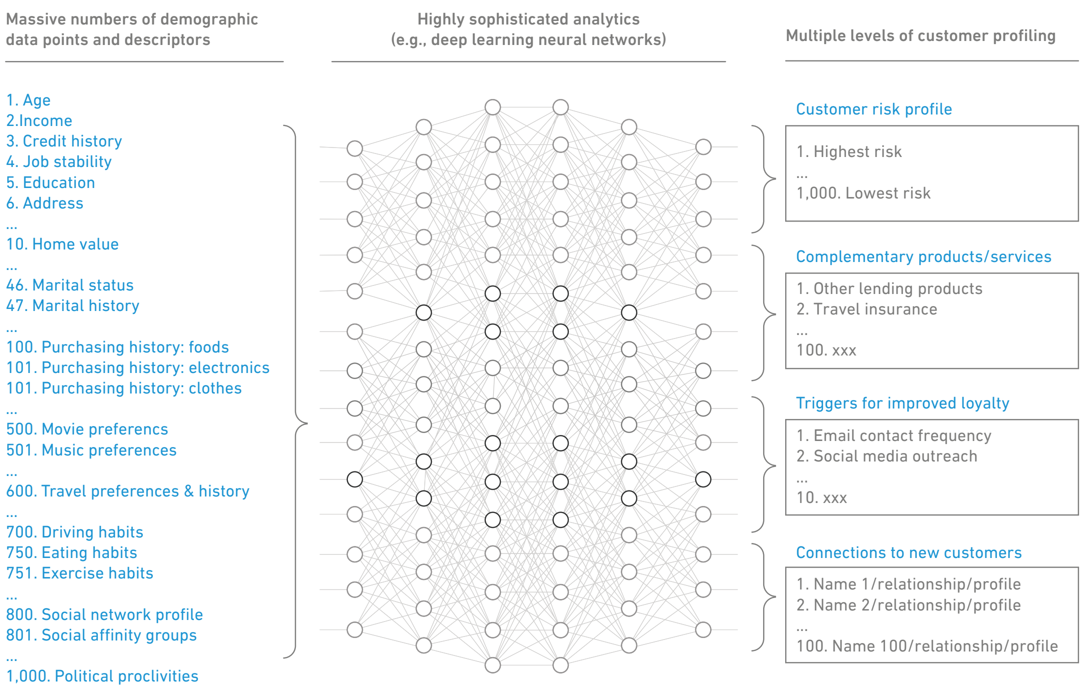
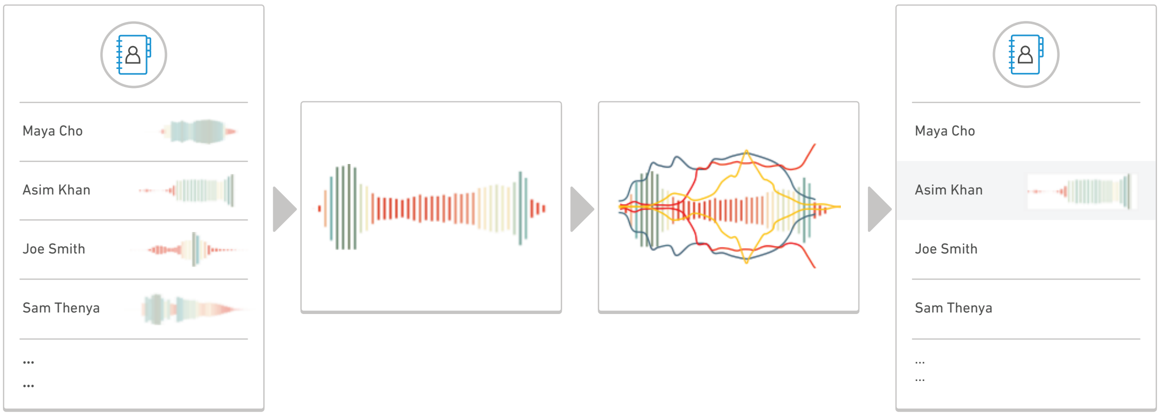


Exhibit 2. By 2010, the amount and type of data began to vastly increase due to the proliferation of smartphones, social media platforms, and other data sources. By using this data and even more powerful analytical tools than those used in previous decades, companies in many industries are able to model consumer behavior and preferences with an astonishing degree of precision.

Another telling illustration of the evolution of analytics and AI is through the transformation of speech recognition and its uses. Exhibit 3 and Exhibit 4 illustrates how a computational problem once at the frontiers of AI research has now become a foundational and commonplace convenience; and once established, how that foundation has now led to an impressive new set of seemingly "intelligent" tools and devices. Speech recognition has been an early AI pursuit since the 1950s and '60s.

By the 1980s and '90s, as it became possible to digitize the analog signals generated by sounds, early practical applications began to emerge. By the turn of the century, it was possible to use voice to dial someone on mobile phones (Exhibit 3). At the time, such early applications—differentiating an individual name from a list of names, across a range of voices, pronunciations and accents—were considered remarkably impressive, and indeed, intelligent.



1. Standard pronounciations of all names in phone's contact list are digitized

2. User says a name, which is then digitized

3. Name spoken by user is matched against all the names in the contact list

4. Closest match is dialed

Exhibit 3. Early-generation AI systems for applications like speech recognition which appeared impressively intelligent at the turn of the century, are considered relatively simple by today's standards. One common application was voice-enabled phone dialing, in which digitized commands and names were matched within a very limited universe.

Today, when smart speakers like Amazon's Alexa® can interpret not just individual words but full sentences or questions and apply those against the virtually unlimited search space that is the internet, the early speech recognition tools no longer appear particularly intelligent, and no one would refer to them as AI.

Exhibit 4 illustrates a simplified algorithm for such a utility; it shows how greater computational power combined with massive amounts of data and references in the cloud can create the impression of intelligent human-machine interaction. (Note that this simple illustration does not necessarily represent algorithms used by Alexa® or any other commercial product, which are likely much more sophisticated.)

The main reason we emphasize the distinction between AI and conventional analytics in current nomenclature is that we fear there has been far too much conflation between the two, in forums focused on human development.

The phrase "big data" is often used anytime any amount of data—big or small—is aggregated for analysis; similarly, the phrases "deep learning" and "AI" are used, even when the statistical optimization tools are now squarely in the realm of conventional analytics. Tautologically, in order to utilize data analytics, a sufficient amount of data must be available.

Put differently, the sophistication of the analytical tool should be appropriate to the underlying richness and volume of data. Just as analytical tools that are too simplistic for the underlying data will lead to erroneous conclusions, so will using tools that are too complex. Indeed, according to data complexity theory, using analytical models with too many degrees of freedom can lead to errors due to over-fitting[8].

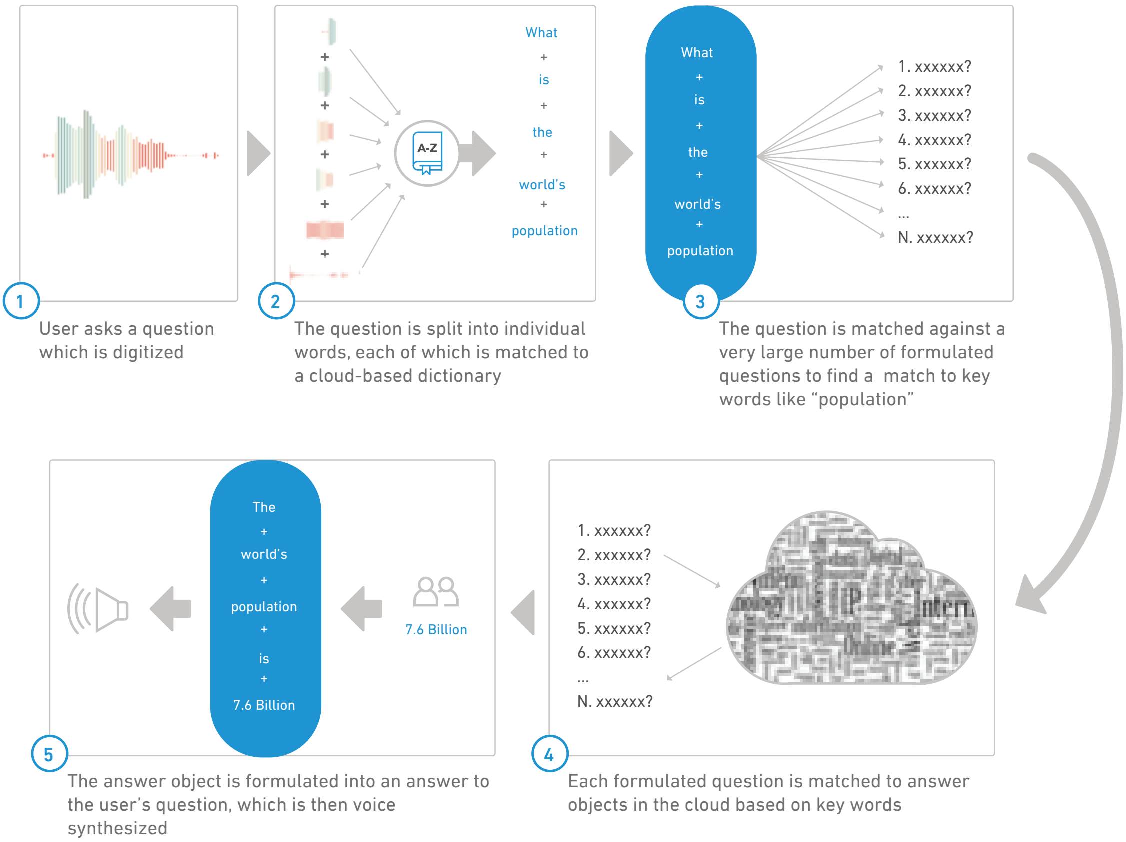


Exhibit 4. New-generation AI tools enable users to have complex interactions with smart speakers which parse sentences and questions, search the web for answers, and formulate responses. As computational power and cloud-based information has grown, so has the ability of machines to perform highly sophisticated tasks which today create the impression of intelligent behavior and interaction.

[8] Data Complexity in Machine Learning", L. Li and Y.S. Abu-Mostafa, Learning Systems Group, California Institute of Technology, May 2006.

3. The Data Density Index and the ability to utilize AI and analytics

A country's ability to use data analytics depends on how much (and what kinds of) data it collects across the public-private spectrum. As such, we introduce the Data Density Index (DDI) as a measure of the strength of a country's data infrastructure and its readiness to utilize data to advance its development objectives. The DDI is a composite of indicators that cover four key areas of digital development:

- Business: How effectively are the country's businesses using ICT and collecting relevant data?
- People: How widespread is ICT use (including social media) among the country's citizens?
- Government: How effectively is the country's government using ICT in its public services?
- Infrastructure: How well-developed is the country's overall ICT infrastructure?

The DDI Business indicator is derived from three metrics in the Networked Readiness Index calculated by the World Economic Forum[9]: ICT use for business-to-business transactions; Internet use for Business-to-consumer transactions; and Impact of ICTs on business models.

The People indicator is derived from the ICT Development Index calculated by the International Telecommunications Union[10], and the Use of Virtual Social Networks metric reported by the World Bank.

The Government indicator is derived from the E-Government Index calculated by the UN Department of Economic and Social Affairs[12].

The Infrastructure indicator is derived from the Telecommunication Infrastructure Index calculated by the UN Department of Economic and Social Affairs[13], and the International Internet bandwidth per user reported by the World Bank[14].

To derive the indicators, each metric was first normalized to a value between 0 and 1, where 1 corresponded to the highest score among all the world's countries[15]. A simple average of all of the metrics comprising each indicator was then calculated for each country. Finally, each indicator was multiplied by 25 and summed for each country (thus giving equal weight to the four indicators), resulting in a maximum DDI score of 100. The DDI was calculated in this way for 136 different countries for which data was available.

---

9 https://widgets.weforum.org/gitr2016/
10 https://www.itu.int/net4/ITU-D/idi/2017/index.html
11 https://tcdata360.worldbank.org/
12 https://publicadministration.un.org/egovkb/en-us/Data/Compare-Countries
13 https://publicadministration.un.org/egovkb/en-us/Data/Compare-Countries
14 https://tcdata360.worldbank.org/
15 Because of the extremely large range of the International Internet bandwidth metric (measured in kilobits per second per user), the logarithm of the scores were calculated prior to normalization.

Together, these indicators are intended to capture the five "V's" often used in the data analytics industry: Volume (are there enough data points captured in the ecosystem?), Variety (is the sample data adequately representative of the range of uses across the broader population?), Value (is the data being captured meaningful or useless?), Veracity (is the data accurate, or are there errors?), and Velocity (is the data current, and being updated with adequate regularity?).

Exhibit 5 shows the calculated DDI scores for 136 countries. The exhibit also demarcates three categories of countries: data deficient, data sufficient and data rich. We define "data deficiency" as having a DDI score of 60 or less. The answer to the "how much data is enough" question is complex and depends on the variance between different population segments across the different data types. The threshold of 60 is a simplification based on a high-level analysis of income/economic population pyramids in developing countries.

Above this minimum threshold and up to a DDI score of 80, we consider a country to be "data sufficient". With the use of appropriate statistical methods, it is possible to extrapolate to the broader population from a sample; therefore, it is not necessary to have complete penetration of data platforms to capture data representing all of the population. As such, we make the (admittedly simplified) assumption that beyond a DDI score of 80, it is likely that more than enough data is captured about a country's population; therefore, we consider such countries to be "data rich".

Not surprisingly, most developing countries are still data deficient despite the recent explosion in smartphones and related apps; this is largely due to the lack of robust data platforms for public services. Governments and other actors can use data from private platforms to measure and improve public services, but it is likely that private companies will only invest in collecting data that will further their profit motives. It is, therefore, in the broader public interest for governments to invest in systems which can be supplemented by private platforms—rather than rely solely on private platforms.

- Despite the rapidly growing rates of smart phone penetration, many sub-Saharan African countries have very low DDI scores making them data deficient, due to limited government commitment to data use and rudimentary ICT infrastructure. While still data deficient, Kenya has a relatively higher DDI score due to its high adoption of smartphone apps for mobile payments—and the derivate tools, services and analytics that are enabled by such apps. Several southern African countries also have relatively high DDI scores, including South Africa, Namibia and Botswana, due to their more developed ICT infrastructure.

- India, like much of the developing world, has experienced significant penetration of smart phones and a range of apps, but remains data deficient. However, India has implemented two major initiatives, the Aadhar digital ID system (which has achieved nationwide penetration, serving over 1.2 billion citizens), and building on that, the data backbone known as India Stack system for digitizing information related to a broad range of services to Aadhar ID-holders. This foundation will serve Indian citizens well in the long run, in holding their government service providers accountable. The India Stack system is still in its early days, and penetration will likely pick up in the years to come—at which point India will likely become data sufficient. The early success and long-term promise of Aadhar and India Stack are encouraging other countries to consider their own versions of such systems.[16]

- China, as a fast-emerging global power in the ICT space has a rapidly growing competence (even dominance) in computing and AI. Its growing IoT infrastructure, massive country-specific platforms for commerce and social media (e.g., Weibo, Alibaba), combined with the government's well-documented interest in closely monitoring citizen activity has made China data sufficient. Its large population and increasing digitization creates major opportunities for AI and big data. There are concerns, however, about the use of AI for ubiquitous state surveillance including the Social Credit System being developed by the Chinese government to assess citizens' reputations.

- OECD countries like the US and many European countries have very high levels of penetration on many public and private platforms, thus are data rich. Some gaps remain, for example, in US health systems which are still largely controlled by private healthcare providers who have limited incentive to share data for external analysis (especially considering the strict privacy protections in place by the Health Insurance Portability and Accountability Act[17]). While aggregated and synthesized data sets are available, it is difficult to extract large raw data sets for deep analytics to find complex patterns.

- Estonia has made bold strides in digitizing the full spectrum of services through its e-Estonia initiative[18], and is considered data rich. It is still developing its economy and infrastructure, thus is well positioned for future growth in extracting and utilizing data.

---

[16] "Should Other Countries Build Their Own India Stack?" A. Raman and G. Chen, CGAP, April 2017.
[17] US Department of Health and Human Services:
https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.
[18] https://e-estonia.com/.

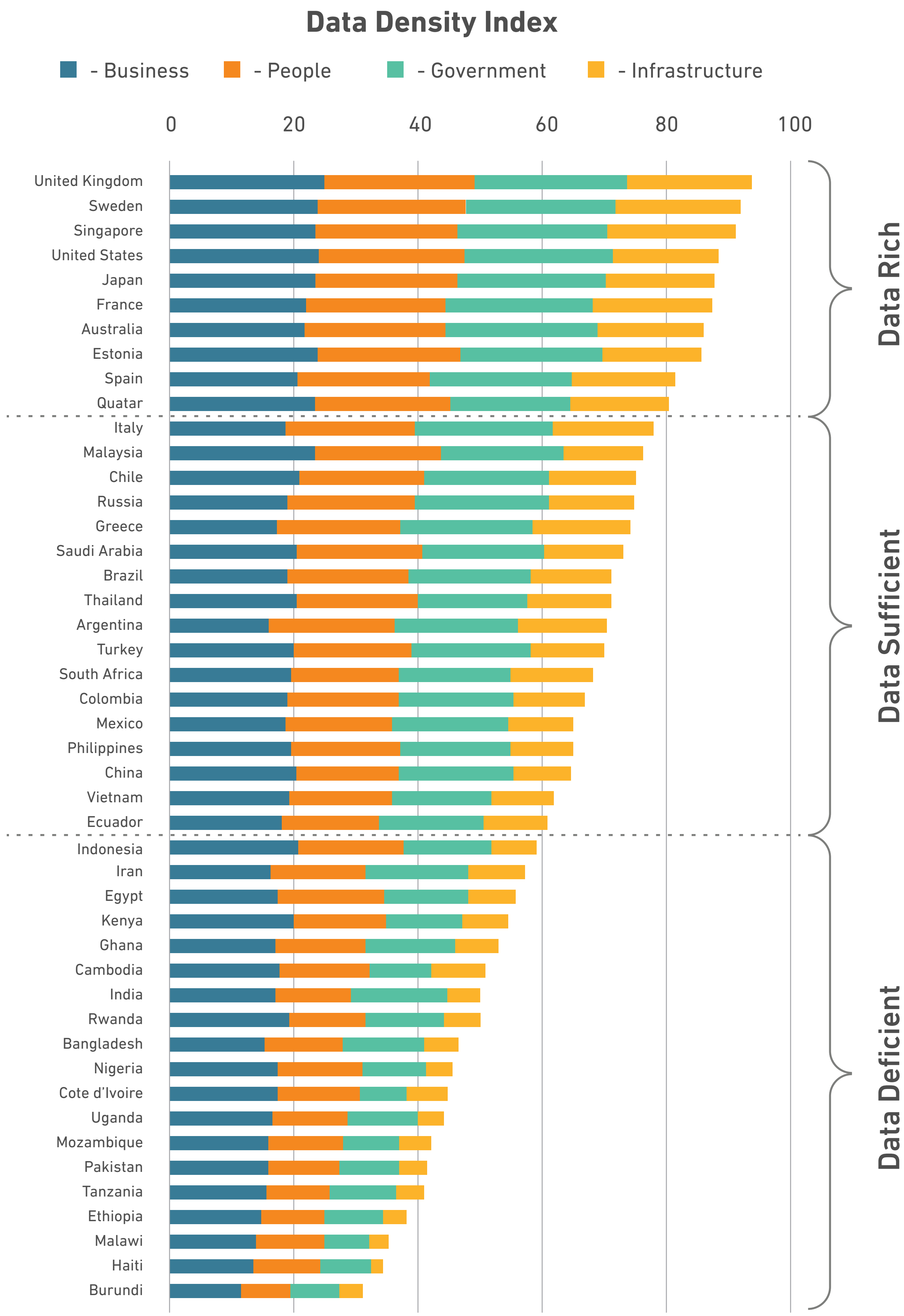


Exhibit 5. The Data Density Index of 136 countries around the world, based on indicators of four key areas of digital development: Business, People, Government, and Infrastructure. Based on these indicators, we categorize countries into three groups: data deficient, data sufficient and data rich. Despite the recent explosion in smartphones and related apps, most developing countries are still data deficient due to the lack of robust data platforms for public services. It is important to note that since a number of metrics comprising the composite DDI are proxies, the index is directional rather than highly precise.

One important outcome of digitization in India is the prospect of virtually unmatched quantities of data, which can create a massive data and learning platform for AI systems. For example, efforts are already underway to develop AI system which use large numbers of digitized retinal scans from India to automatically detect diabetic retinopathy[19]. In principle, such a system developed with data from India can be used in countries around the world.

Exhibit 6 shows the relationship between the Data Density Index and the Human Development Index (HDI) calculated by the United Nations Development Programme[20]. As expected, there is a strong correlation (r^2 = 0.91) between the two indices. Nevertheless, some countries have DDI scores that are significantly higher or lower than would be expected based on their HDI scores. For example, Bahrain, UAE, South Africa, Kenya and Rwanda have higher-than-expected DDI scores. Other countries have lower-than-expected DDI scores, including Iran, Venezuela, Algeria, Sri Lanka and Gabon.

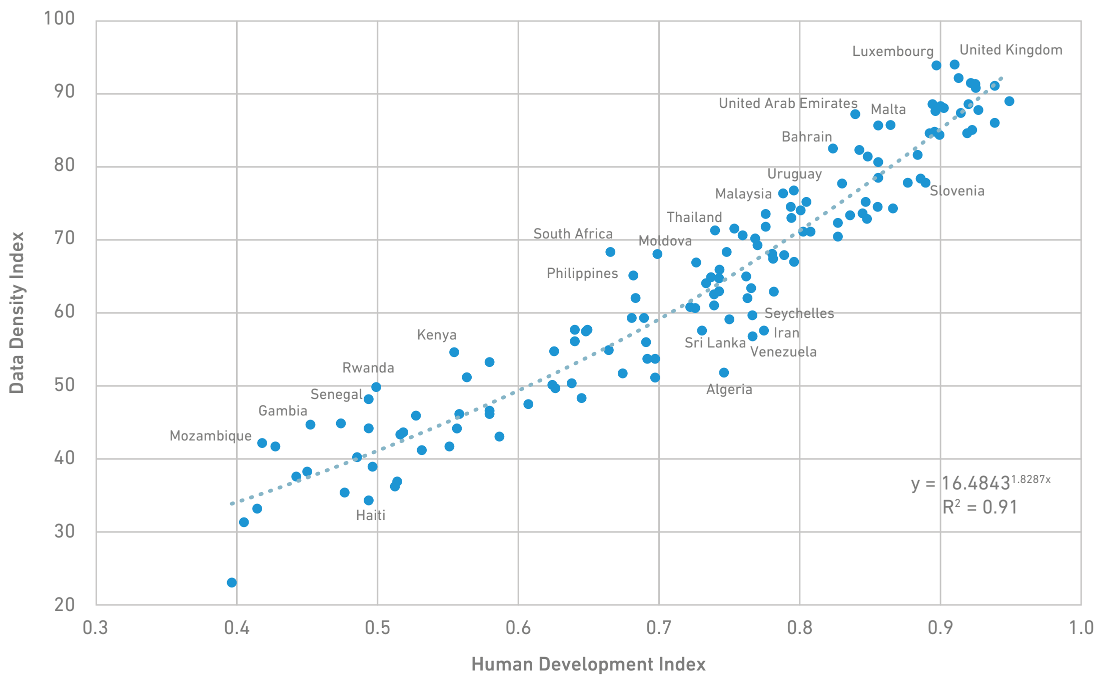


Exhibit 6. There is a strong correlation (r^2 = 0.91) between the Data Density Index and the Human Development Index. On this figure, countries are labeled if their DDI scores are at least 6 points higher or lower than their predicted DDI score based on exponential fit.

[19] "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence", R. Rajalakshmi, et al., Nature, March 2018.
[20] http://hdr.undp.org/en/data

4. The most important levers for human development, and their dependence on data/AI

The following analysis considers the value of AI and data-driven solutions for four areas critical to human development: food security and agricultural development, health, energy access, and education.[21] For each of these issues, we identify the most impactful interventions (based on our 50 Breakthroughs study, an in-depth search of the literature, analysis of the strategies employed by leading institutions in global development, and interviews with experts).

We then categorize these interventions along two dimensions:

- Dependence on data analytics and AI: Can the solution be developed/implemented with (i) limited need for data, (ii) with conventional analytics, the likes of which have been in use in industrialized contexts for the past 2-3 decades, or (iii) with "big data" and new-generation AI tools. The following discussion shows that many of the interventions have limited need for sophisticated data analytics; some can really benefit from sophisticated—but conventional—analytical tools, and a small number require AI as it is currently recognized.

- Is the intervention (i) direct (i.e., is it a primary driver of the satisfaction of human needs), (ii) second-order (i.e., is it an enabler of a direct intervention, or does it rely on a direct intervention for delivering impact) or (iii) tertiary (i.e., is it an enabler of a second-order intervention, or otherwise helpful for the broader ecosystem, but not directly connected to primary drivers of impact).

  The main purpose of this categorization is to underscore the argument that while data-driven insights can lead to significant improvements in decisions, strategies and policies (and in some cases, automation can decrease the need for skilled human capital), there are often more fundamental challenges that needed to be overcome first. When the primary limiting factors are physical goods (e.g., fertilizer, water for irrigation) and infrastructure (e.g., clinics, solar mini-grids or toilets), there will be strict limits to the effectiveness of data, insights and automation to make real improvements in the lives of the poor.

  Also, insights can lead to impact only with effective on-the-ground execution. In other words, second-order and tertiary interventions, even when necessary, are not sufficient for impact without effective direct interventions. For example, remote-sensed knowledge of illegal logging or fishing is mere documentation of abuse, and cannot stop such activities without effective on-the-ground enforcement.

[21] While this is admittedly a small subset of the full range of SDG topics, we believe it is enough to illustrate our core thesis about the role of data and AI for human development.

4a. Food security and agricultural development

Food security and agricultural development faces a two-pronged challenge: (i) in Sub-Saharan Africa and parts of South Asia which did not benefit from the Green Revolution of the 1960s and '70s, the imperative is to significantly increase crop yields and livestock health/productivity, and to get them to market; (ii) in parts of the developing world (particularly South/Southeast Asia) where agronomic practices have matured due to the Green Revolution and its derivative effects, the challenge is to significantly improve the environmental sustainability of food production.

The key hurdles that need to be overcome for smallholder farming[22] in the agriculturally underdeveloped parts of Sub-Saharan Africa and South Asia are:

- The vast majority of smallholder farmers lack access to irrigation, fertilizers and inputs for pest control; this is compounded by a lack of smallholder finance for purchasing these inputs.

- Support systems for livestock farming are weak, at best, for breeding, vaccination and veterinary services; this exacerbates the problem of insufficient nutritious fodder for the animals.

- There is increasing vulnerability to rising temperatures and extreme weather events such as floods and droughts; this is exacerbated by climate change and land cover change such as urbanization and deforestation, as well as biotic stresses (pests, pathogens) and erosion of fertile soil.

- Local capacity for processing agricultural produce is weak or nonexistent, as a result of which much of the economic value from agricultural commodities gets transferred to countries which have processing capabilities. In addition to the high cost of processing equipment, a major underlying hurdle is the severe lack of financing for small and medium-sized enterprises (SMEs) in the agricultural processing business.

- Farmers have limited direct access to markets, for staples as well as cash crops. The former leads them to need to store a full year's worth of product; and the latter deprives them of the ability to increase incomes.

- The agronomic practices used by the majority of smallholder farmers are sub-optimal, which exacerbate the already difficult challenges they face due to a lack of access to critical farming inputs.

In the parts of Asia which benefited from the Green Revolution and have relatively mature agricultural systems, there are major challenges related to the environmental damage from farming, including groundwater depletion, water pollution from fertilizer runoff and soil erosion.

[22] It is important to note that there is an active debate about whether smallholder farming is the appropriate model for improving food security, or for alleviating poverty. We do not have a strong perspective on that debate, and are drawing from the current strategies of the major institutions focused on the topic.

There are a number of direct interventions which can directly address some of these challenges listed above. None of these interventions have much need for data-driven analytics. (There may, of course, be a need for appropriate ICT tools and reliable data.)

1. A suite of affordable on-farm implements for irrigation (solar pumps and drilling mechanisms for shallow groundwater) and weed removal.

2. Fertilizer, either through reliable distribution of conventional fertilizers, or via new mechanisms to produce fertilizer (e.g., from small-scale alternatives to the Haber-Bosch process, or efficient composting of organic waste).

3. A steady supply of livestock vaccines for the most destructive diseases and pathogens (endoparasites, *peste des petits ruminants*, Contagious Bovine Pleuropneumonia, etc.). Such an intervention can be bolstered by a cold chain for artificial insemination. Both of these, ideally, will be part of extension workers' kits.

4. High-nutrient and low cost, sustainable animal fodder.

In addition, there are a number of critical second-order interventions which will improve access to, or application of the primary interventions described above:

5

Improved access to finance for farmers, as well as other actors along the value chain such as local food processors. More than anything else, this requires an aggregation of funds, and effective mechanisms to assess risk, lend and collect the loans. Accurate risk assessment requires robust analytics, much like credit agencies in the industrialized world employ. However, conventional modeling tools—in use for at least two decades in industrialized markets—will suffice. The volume and type of data currently available for smallholder farmers and small/medium agribusinesses is most appropriate for conventional tools rather than AI.

6

Crop [micro-]insurance, particularly against catastrophic losses related to weather and pathogens. Insurance is a complex business, requiring adequate reserve funds, good risk assessment and pricing, collections, and potentially a pool of funds for reinsurance. As with lending, insurance requires robust data-driven models for accurate risk assessment for insurance (at the individual policyholder level, as well as for customer segments, other actuarial analyses and climate/weather modeling). In this case too, the volume and type of data currently available is most appropriate for conventional tools rather than AI.

7

Dynamic information on markets connecting farmers to potential buyers, with the latest market prices and other information to disintermediate the value chain and optimize transactions. As in the cases of lending and insurance, conventional analytics combined with existing ICT tools can go a long way in building very effective solutions.

8

Affordable (ideally energy-efficient) processing equipment, especially for key cash crops so that the economic value of the commodities remains local. There is limited need for data-driven analytics in such processing.

9

Tools and equipment for reducing post-harvest losses, including dry storage (especially for grains) and refrigeration for perishables.

There are also a small number of tertiary interventions:

10

Interactive, dynamic content for farmers and processors, customized to their hyper-local or even individual needs and context, delivered via social media and other engaging means. Such a tool can require sophisticated software tools similar to what is currently used in many social media outreach applications. However, there is limited need for anything resembling AI.

11

Highly granular data on yields, soil health, pathogens, etc., to improve policy-making by the relevant institutions. This will require robust tools for data collection, organizing and analysis; however, there is limited need for AI.

12

A new generation of genetically engineered seeds, particularly for drought and heat tolerance. CRISPR has made genetic engineering of seeds significantly more feasible and cost-effective. While transgenetic GMOs (i.e., introducing DNA from unrelated species) can be problematic, cisgenetic hybridization (i.e., introducing DNA from the same or related species) may be much more predictable and acceptable. Among the many problematic aspects of genetically engineered seeds has been that their interactions with the broader ecosystem can be highly unpredictable, especially over time. While it may be possible to model the immediate impact of the interaction between a new seed variety in its environment, the second-, third- and higher-order effects can be very difficult to predict.[23] In order to address this hurdle, it may be possible to develop highly sophisticated AI simulations to model multilateral and combinatorial interactions. However, there is currently very limited data on which to build such models. Also, given the highly complex nature of interactions between the myriad of organisms in any given ecosystem, it will be very difficult to ensure that the AI models are accurate.

[23] Note the example of BT cotton in India and China, where cotton susceptible to bollworms was genetically engineered to integrate the DNA of the bacterium *Bacillus thuringiensis* (BT), which produces toxins harmful to the bollworm. Initial rollout of BT cotton found that while bollworm infestation was successfully overcome, it opened the door to secondary—equally destructive—pests which the bollworms were keeping out. While the problem of secondary pests was eventually resolved for BT cotton, it does underscore the imperative to understand second-, third-, and higher-order impacts of genetically engineered organisms.

Exhibit 7 summarizes the above discussion along the two assessment dimensions: dependency on data/AI tools, and type/level of intervention. As the summary shows, 5 of the 12 interventions can benefit from robust, conventional data analytics currently limited to commercial operations in industrialized contexts. Only one—multigenerational modeling of complex environmental interactions due to introduction of genetically engineered seeds—can be improved using AI tools. However, it is far from clear that such models will solve the underlying problem.

Type of Intervention	Limited dependence on data-driven analytics	Conventional analytics and decision-making tools	AI and big data
Tertiary enablers		11. Granular data for improved policy-making 10. Training content (customized generation and dissemination) on optimal agronomic practices	12. Improved seeds for drought and heat, with robust models for understanding impacts on ecosystems
Second-order enablers	9. Post-harvest loss reduction including dry storage and refrigeration 8. Processing equipment for key cash crops	7. Market information: pricing and match-making between sellers and buyers 6. Crop [micro-] insurance 5. Funds for financing for smallholder farmers, agricultural processors and other SMEs in the agricultural sector	
Direct Interventions	4. High-nutient animal fodder 3. Lifestock vaccines for key diseases/ pathogens, and cold-chain-enabled artificial insemination for improved breeding 2. Steady supply of fertilizer (conventional or via alternative mechanisms e.g. small-scale Haber-Bosch process or efficent composting) 1. Affordable on-farm implements for irrigation and weed removal		

Dependence on big data/AI

Exhibit 7. Of the most important interventions required for food security and agricultural development, five require robust conventional analytical solutions. Only one solution (# 12 , at the top-right of the matrix) requires new-generation AI; however, even that is highly speculative, given the underlying complexity of the problem.

4b. Health

By almost every measure of health and wellbeing, low income countries suffer a disproportionate share of the world's mortality and morbidity burden. The challenges that need to be overcome are many and complex, and the following is an admittedly simplified summary of those challenges:

- There is a grossly inadequate infrastructure for delivering primary care: insufficient numbers of clinics, and too few trained nurses and physicians. There is a heavy reliance on untrained, uncertified community health workers.
- A large number of births occur at home or in inadequate clinical facilities, administered by under-trained and under-equipped clinicians. This contributes to a high incidence of maternal and neonatal mortality.
- The tropical climates common to most developing countries lend themselves to the presence of an enormous range of pathogens and vectors, which exposes their populations to many more infections—ranging from entrenched diseases like malaria to epidemics like ebola—than their counterparts living in higher latitudes and colder climates.
- Most low-income countries lack an adequate sanitation infrastructure, as result of which diarrheal disease is widespread among children. Frequent bouts of diarrheal disease can lead to chronic malnutrition (especially considering the dearth of nutrient-dense foods) and compromised immune systems.
- Widespread malnutrition and weak immune systems, combined with the exposure to pathogens, lead to high incidence of major infectious diseases such as TB, malaria, pneumonia and diarrheal diseases. The weak healthcare infrastructure and limited means for diagnosing and treating diseases leads to higher morbidity/mortality as well as transmission.
- Most low-income countries do not have an adequate health insurance system as a backdrop for either preventive or emergency care.

- A number of pathogens (e.g., for TB and malaria) are developing resistance to the current generation of drugs.
- While there has been some progress in fighting infectious diseases, there has been a dramatic increase in the incidence of noncommunicable diseases like diabetes, hypertension, heart disease and cancers. In addition, there has been a recognition that mental health is an important issue that has not received enough attention.
- A growing number of deaths are caused by pollution associated with modern industrial development, such as vehicle exhaust, industrial emissions and chemical releases. Deaths due to modern pollution are overtaking deaths due to traditional pollutants including household air pollution, unsafe water sources and inadequate sanitation.
- Most medical devices on the market—for both diagnostics and treatment—are too expensive for low-income settings. Many of them also require reliable electricity as well as high levels of technical training to operate. This makes them unaffordable and unusable in remote, low-income clinical settings.
- There are limited financial incentives for the private sector to invest in diagnosing and curing diseases among low-income populations. As a result, there is an utter lack of R&D in developing the necessary pharmaceuticals and medical devices. Many existing pharmaceuticals (particularly vaccines) are highly temperature-sensitive; the absence of reliable cold chains significantly hampers last-mile distribution.

More than any other aspect of human development, access to quality, affordable healthcare can benefit from technology advances. The following are key direct interventions to address the above challenges; while these may require extraordinarily complex R&D, they likely do not need particularly complex data-driven analytics or AI.

1

Large primary care networks, either private or via public-private partnerships (PPPs), which can put in place standards of quality and service, utilize economies of scale, and have broad reach. In a number of countries, PPPs are being explored as viable, accountable alternatives to a traditional government system.

2

A new generation of effective pharmaceuticals, particularly:
a. Vaccines for HIV/AIDS, Malaria and TB;
b. Single dose or short course TB antibacterials which can improve regimen adherence and curb drug resistance;
c. Microbicides to protect against HIV/AIDS and HPV;
d. A complete cure for malaria that eliminates it from the human body.

3

Nutrient-dense, culturally appropriate foods for infants can be an effective tool to tackle malnutrition in the critical early years of life.

4

Long-lasting, non-chemical, low cost spatial repellents for mosquitoes and other vectors.

5

A reliable cold chain for vaccine delivery, especially to remote areas. In the long run, thermostable vaccines—currently still in early R&D phase—are likely the best solution.

6

Toilets which destroy pathogens locally, without the need for a sewer infrastructure (e.g., composting toilets).

Reliable systems to deliver care, especially at the primary level, are perhaps the most critical intervention to ensure that the new generation of pharmaceuticals and diagnostics technologies lead to improved health outcomes. To that end, several important second-order interventions are required, including:

7

A suite of point-of-care diagnostic devices for essential urine, blood and vitals tests is necessary for providing primary care. These should be simple enough to use by nurses (and potentially even community health workers), require limited calibration and maintenance, and capture data digitally. A number of affordable new devices are beginning to appear on the market, but accuracy and robustness remains a challenge. Data-driven monitoring mechanisms across large numbers of sites can significantly improve quality control and performance of such devices; however, this does not require AI-based tools.

8

AI-based automated diagnostics using existing technology: Many types of diagnostic data require trained clinical staff. Given the large skill gap, AI-based tools can be tremendously helpful for automated detection of diseases/conditions from radiological images, retinal scans and other digitized data. If enough images are available with expert input on which diagnostic features are linked to which diseases and conditions, the current generation of AI techniques should be adequate.

9

Rules-based clinical protocols: The proliferation of tablets and smartphones now makes it highly feasible to incorporate standardized clinical protocols for patient care, based on the patient's demographics, medical history, diagnostics and the clinician's skill level (physician vs. nurse vs. community health worker). Such a rule-based system can be constructed using conventional analytical tools; indeed, ITT has deployed such a system with strong early results.

10

Health insurance has proven to be a critical component of meaningful healthcare systems, wherever it has been implemented. Among developing countries, Kenya appears to be having strong early success with its National Hospital Insurance Fund. All evidence suggests that low-income populations can be served at a large scale only through government-funded programs like NHIF. Such government-funded programs ideally do not price policies on the basis of a policyholder's risk profile; hence, they do not necessarily need sophisticated analytical modeling. Even private insurance programs can be implemented with existing analytical tools.

11

Models for tracking early indicators of outbreaks of diseases like Ebola and rapidly forecasting patterns for transmission and spread. Advanced data-driven and AI tools can make a powerful contribution to such solutions.

12

A new generation of AI-based diagnostic techniques, bypassing conventional biomarkers: Even as effective diagnostics are developed for basic conditions, there is a need for a new generation of integrated diagnostic platforms for more complex and difficult-to-diagnose conditions. One approach is to attempt point-of-care multiplex diagnostic immunoassays which use different types of samples (e.g., blood, urine, sputum), and test for multiple biomarkers (and diseases) from a single patient interaction. Such an approach can be extraordinarily complex, requiring potentially decades of R&D. AI-based tools may provide a breakthrough alternative. It is possible that digitized images (e.g., x-rays, ultrasounds, retinal scans) contain heretofore unknown indicators of a broad range of diseases and conditions. As the database of such images and corresponding diseases/conditions grows, the latest AI techniques may be able to find the link between the indicators and their clinical implications. Such an approach is unproven, but has many experts excited about the possibilities.

A small number of tertiary interventions can also help improve the effectiveness of the above interventions.

13

Customized training tools: The dramatic proliferation of smart devices in recent years has laid a strong platform for highly customized skills assessment and training of clinicians. Making such tools an essential component of healthcare delivery systems—public and private alike—can go a long way in addressing the current skills gaps. Such tools can be developed using conventional data-driven customization and content generation tools.

14

Granular data for policymaking: A key component of effective policymaking is robust data with high spatial and temporal resolution. The volume and quality of data has been steadily increasing in recent years, and the stage is set for a major step-change over the next decade, using existing data collection and analytical tools.

Exhibit 8 summarizes the key interventions in improving health and wellbeing in developing countries. As the exhibit summary shows, data-driven analytics and the current generation of AI can make a significant contribution towards a number of solutions. New-generation AI tools offer the exciting possibility of a new paradigm for diagnostics, potentially bypassing conventional biomarkers.

Type of Intervention	Limited dependence on data-driven analytics	Conventional analytics and decision-making tools	AI and big data
Tertiary enablers		14. Granular data for improved policy-making, incorporating local health data and trends 13. Customized training and certification tools for clinicians	
Second-order enablers	7. Suite of point-of-care primary care diagnostic devices	10. Large-scale (ideally government-run) health insurance programs 9. Rules-based clinical protocols based on patient demographics, medical history, point-of-care diagnostics, etc. 8. Diagnostics for image-based data (x-ray, retinal scans, etc) using existing AI-based automation	12. New generation of AI-based diagnostic techniques, bypassing conventional biomarkers 11. Epidemiological/outbreak tracking and forecasting
Direct Interventions	6. Decentralized toilets (e.g. composting) 5. Reliable vaccine cold chain equipment; or thermostable vaccines in the long run 4. Long-lasting, non-chemical, low cost spatial repellents for mosquitoes and other vectors 3. Nutient-dense, culturally appropriate infact foods 2. New generation of pharmaceuticals: vaccines for HIV/AIDS, malaria and TB; single dose or short course TB antibacterials; etc 1. Large primary care networks (PPP or private/branded) which scale economics, quality standards and broad/deep reach		

Dependence on big data/AI

Exhibit 8. Of the most important interventions required for healthcare, five require robust conventional analytical solutions and existing AI tools. New-generation AI tools can help create a new paradigm of diagnostics and predict disease outbreaks.

4c. Energy Access

Affordable, reliable, clean energy is a key driver of economic development. Low-income communities in developing countries lack access to several forms of energy: electricity for powering household and commercial appliances, fuels for cooking (and in some climates, heating), and energy for powering automobiles.

Roughly 1 billion people around the world lack access to electricity (mostly in rural areas of Sub-Saharan Africa and South Asia); this number can potentially grow if population growth outpaces electrification rates. The key challenges that impede universal access to electricity are:

- Conventional grid infrastructures—with centralized generation and widespread distribution—are too capital intensive for the governments of low income countries. Recognizing the environmental cost of fossil fuels, many governments have begun shifting away from conventional electrification models.

- Renewables, especially solar, are experiencing rapid growth[24]. In recent years solar home systems and, to a smaller extent, mini-grids have emerged as viable mechanisms for electrifying communities without reliable electricity. However, home systems are not designed to power productive electricity uses critical to economic growth, and both home systems and mini-grids are considerably more expensive than conventional grids powered by fossil fuels.

  While the cost of solar photovoltaic panels continues to drop significantly, a purely solar electricity infrastructure remains expensive because of the cost of the rest of the system—power electronics and, in particular, energy storage. Conventional lead-acid batteries, the default storage solution, do not perform well under normal stresses; and lithium-ion batteries are still too expensive.

  Note that, to date, the only financially viable storage solutions (lead-acid and lithium-ion) exist because of their massive market in the automobile industry; battery chemistries designed exclusively for non-vehicle electricity storage have struggled to survive financially. Even if a new battery is successfully developed and commercialized for storing electricity for grid/mini-grid distribution, it is much more likely to find its commercial footing first in industrialized markets before reaching low-income ones.

---

[24] Other renewables (e.g., wind, hydro, geothermal) are highly location specific, and can therefore incur high distribution costs. Geothermal and large hydro projects are also highly capital intensive.

- Energy-efficient appliances, especially mechanized ones using efficient brushless DC motors, are much more expensive than those using AC motors. This increases the consumption and cost of productive, income-generating uses of electricity.

Low-income households in most developing countries use traditional cookstoves with fuels like wood, charcoal or other biomass, which emit high levels of smoke (and in the case of wood and charcoal, also lead to deforestation). This pollution causes respiratory infections, heart disease and lung cancer, leading to over 3 million deaths each year. Over the past decades, a number of fuel efficient, clean, and affordable cookstoves have been developed, but virtually none have achieved sustainable scale.

The main reasons are:

- Cooking food requires a tremendous amount of energy, relative to other basic household energy uses. Furthermore, cooking fuel is required daily, without fail, as every household needs to cook.
- Clean stoves compete against an alternative which is often free (even if time and effort is involved in collecting fuels), and the long-standing traditional choice.
- While there is clearly a health imperative in reducing household air pollution from traditional stoves, it is not clear that there is demand for this. Presumably, this is because those most exposed to high household air pollution may not see any short-term value in cleaner stoves, especially considering their economic means.

Automobiles have historically been too expensive to be a realistic option for low-income household use.[25] On the other hand, affordable vehicles have been a critical enabler of income and productivity, and the recent growth in electric-powered vehicles (e.g., e-rickshaws in India) has been a welcome alternative to gasoline and diesel. However, such vehicles are currently largely powered by coal-fired electricity plants charging lead-acid batteries, with a poor track record on environmental pollution. Solar photovoltaic collectors and lithium-ion batteries are still too expensive for lower-income users.

[25] ITT's 50 Breakthroughs study identifies affordable, efficient, electric household vehicles as one of the most critical cross-cutting technology breakthroughs.

The following direct interventions are key to addressing the various energy access challenges described above:

1

Large (nationwide or province-wide, depending on the population size) solar utilities based on public-private partnerships and capital subsidies (similar to those for conventional fossil-fuel based infrastructures). These utilities can deploy modular networks of solar mini-grids which, over time, can grow with demand and connect to larger grids, eventually forming massive distributed generation infrastructures.

2

Productized, modular "drop-in-place" solar mini-grid systems which can significantly reduce installation cost and time. Such systems should also have smart-metering with mobile payments and a variety of payment structures (pre-pay, pay-as-you-go, post-pay). Over the past year or two, a number of companies (including ITT) have been developing such products, although none has reached scale as of 2018.

3

A storage solution for solar mini-grids, which leverages existing supply chains of automobile batteries. Two possibilities are emerging: second-life lithium-ion batteries (which automobile manufacturers tend to replace once they reach 70-80% of original capacity), and an advanced lead-acid chemistry which addresses many of the challenges faced by conventional lead-acid batteries. As of Q4 2018 neither battery has demonstrated its business case.[26] Refurbished lithium-ion car batteries can also be a valuable storage solution for low-cost electric vehicles like e-rickshaws, assuming there is a large enough supply of used/refurbished batteries.

[26] Over the past two years, ITT has been working with Tata Power in India to test the technical performance of a number of battery chemistries. "Emerging storage solutions for solar mini-grids: Results from tests of high-potential batteries," ITT, May 2018.

4

Commercially viable clean cookstoves and/or cooking fuels which accommodate the variety of local cooking utensils and practices, providing affordable, reliable and sustainable options for household cooking.

One key second-order intervention can help support the primary interventions described above:

5

A new generation of affordable energy-efficient appliances, especially for productive uses which can drive growth of small enterprises (e.g., agricultural processing, refrigeration, woodwork). This intervention does not depend on advanced analytics or AI solutions.

There is also a tertiary intervention which can support various interventions described above:

6

Accurate models and data on likely energy usage in communities based on population size and projections of consumption patterns and economic growth, so that utilities can plan optimal capacity for their systems, and regulators can plan optimal policies. Such tools can also be used for smart grid management, by matching electricity demand and supply. Conventional analytic techniques are sufficient for these tasks, which do not require data-intensive AI.

Type of Intervention	Limited dependence on data-driven analytics	Conventional analytics and decision-making tools	AI and big data
Tertiary enablers		6. Data and predictive models on community energy usage, so that utilities can plan capacity and regulators can develop appropriate policies	
Second-order enablers	5. New generation of affordable energy-efficient appliances, especially for mechanized productive uses		
Direct Interventions	4. Commercially viable clean cookstoves and/or cooking fuels 3. Storage solutions for solar mini-grids, which leverage supply chains for automobile batteries 2. Productized, modular “drop-in-place” solar mini-grid systems which can significantly reduce installation cost and time 1. Large solar utilities based on public-private partnerships and capital subsidies (similar to those for conventional fossil-fuel based infrastructures)		

Dependence on big data/AI

Exhibit 9. In energy access, data analytics can play a targeted, indirect role.

4d. Education

Since the Millennium Development Goals were announced at the turn of the century, some aspects of education have witnessed significant improvement: enrollment in primary schools; offerings in technical/vocational education and training (TVET) with an emphasis on entrepreneurship; and a recognition that investing in early childhood development is critical to longer-term cognitive development. However, a number of significant challenges still need to be overcome, before universal quality education in low-income countries becomes a reality, and leads to broader economic growth and equitable participation.

- Despite the improved enrollment rates in primary school, tens of millions of children of primary age still remain unenrolled. There are not enough schools or teachers, and per-student spending in low-income regions continues to be a small fraction of that in higher income regions.

- Improvements in enrollment numbers have not been accompanied by an equal rise in quality. The key challenge has been a continued dearth of well-trained teachers supported by ongoing training and appropriate tools, exacerbated by high student-teacher ratios.

- Pressures to generate income continue to force children to drop out of school once they reach their teens, leading to a significant drop-off in enrollment rates in secondary school.

- Traditional beliefs on the role of women and girls in society—although gradually becoming less significant—continue to discourage girls' education, especially as they age.

- There continues to be a disconnect between higher education curricula and the most promising job opportunities, as a result of which too many corporate roles in African countries are filled by expatriates from other countries.

To achieve universal, quality education, there are few alternatives to long-term investment in effective public education systems: schools, teachers, quality control mechanisms, linkages to employment, and policies to incentivize parents and teachers to make long-term education outcomes a priority. However, a small number of potentially disruptive interventions can pave the way.

One emerging direct intervention is:

1

Large networks of branded, low-cost private schools at all education levels: primary, secondary and vocational. There is strong evidence suggesting high demand for affordable, accountable education among low-income communities, especially in urban areas. Large, branded networks can take advantage of scale to develop standards of excellence as well as platforms for collecting data on student and teacher performance. For higher age groups, large, respected vocational training schools can create pipelines for job and entrepreneurship opportunities.

There are a number of important second-order interventions to improve education access and quality:

2

Incentives for families (e.g., conditional cash transfers) to invest in children's education (a la Brazil's Bolsa Familia program[27]). Robust data and analytics can be very useful in designing optimal programs and incentive packages.

3

Incentive programs for schools and teachers, closely linked to educational outcomes, investment in teacher training, and other improvements in quality and access. As with incentives for families, robust data and analytics can be very useful in designing optimal programs and incentive packages.

[27] "Brazil's Antipoverty Breakthrough: The Surprising Success of Bolsa Família", J. Tepperman, Foreign Affairs, January 2016.

4

Student loan programs, especially for vocational training and entrepreneurship. Reliable data and analytics can help understand the success rates of various training programs, and shape loan and incentive programs accordingly.

5

Intelligent teaching systems, enabled by semi-automated content curation, automated grading to dynamically adjust learning speeds and remedial vs advanced content, and interfaces/ functionalities for both students and teachers. Over the past 5 years, there has been an explosion of increasingly powerful learning and teaching tools which combine cloud-based and offline capabilities as well as curated content. A small number of additional functionalities can lead to a true step change, creating intelligent systems that can transform both learning and teaching experiences.

These include:

a. Automated grading to gauge student competence. While it is easy to grade students on multiple choice or simple text answers, it is much harder to assess the quality of more involved, long-form answers.

b. Automated curation of supplemental (e.g., remedial) content from the cloud, based on the automated grading mechanism described above.

c. Translation of curated content into local languages.

Aspects of all three functionalities exist, and it is likely only a matter of a few more years before this next generation of intelligent learning systems becomes a reality.

Type of Intervention	Limited dependence on data-driven analytics	Conventional analytics and decision-making tools	AI and big data
Tertiary enablers			
Second-order enablers		4. Student loan programs, especially for vocational training and entrepreneurship 3. Incentive programs for schools and teachers, linked to improvements in quality, access and educational outcomes 2. Conditional cash transfers and other incentive programs for families to invest in education	5. Intelligent teaching systems, enabled by semi-automated content curation, automated grading to dynamically adjust learning speeds and remedial vs advanced content and interfaces/functionalities for both students and teachers
Direct Interventions	1. Large networks of branded, low-cost private schools at all education levels: primary, secondary and vocational		

Dependence on big data/AI

Exhibit 10. There are a number of important traditional interventions to improve education, such as investing in more schools and teachers. Interesting additional interventions include well-designed incentive and support programs which can benefit from conventional analytics. In addition, a new generation of AI-driven learning tools can make a significant contribution.

5. The risks associated with widely deployed data acquisition platforms and AI

The recent landmark Toronto Declaration on human rights in the context of machine learning begins with the following preamble[28]:

"As machine learning systems advance in capability and increase in use, we must examine the positive and negative implications of these technologies. We acknowledge the potential for these technologies to be used for good and to promote human rights but also the potential to intentionally or inadvertently discriminate against individuals or groups of people. We must keep our focus on how these technologies will affect individual human beings and human rights. In a world of machine learning systems, who will bear accountability for harming human rights?"

The implication of the question ending the opening of this preamble is that the technical mechanisms required to ensure specific groups are not targeted are unclear, even if such protections are acknowledged in more abstract human rights agreements.

A number of recent high-profile scandals[29] related to big data and AI have sparked concerns, but appear to have led to a general acceptance of the risks associated with the technologies, even in industrialized countries with ostensibly robust policy ecosystems and legal protections[30]. As potentially problematic as these risks are in countries with robust legal protections, the challenges are likely to be significantly higher in countries without such protections.

i

Notwithstanding the potential benefits of data analytics and AI, there are a number of legitimate concerns about government surveillance of citizens, targeting of specific groups (based on religion, ethnicity, political proclivities, sexual orientation, or other oft-discriminated-against traits), and other forms of exploitation and persecution.

---

[28] "The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems", Access Now, May 2018.
[29] "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens", The New York Times, March 2018.
[30] "The Risks of Artificial Intelligence to Security and the Future of Work", Rand Corporation, 2017.

There is strong evidence to suggest that lower income countries tend to have weaker governance and rights protections (Exhibit 11). Indeed, today, even without robust data infrastructure, governments of the poorest countries can use any of a number of commercially available tools to surveil the communications of their citizens. While there are a number of forums dedicated to these concerns and the necessary protections, it is unclear when appropriate legal frameworks will be put in place and whether such protections can effectively eliminate abuse.

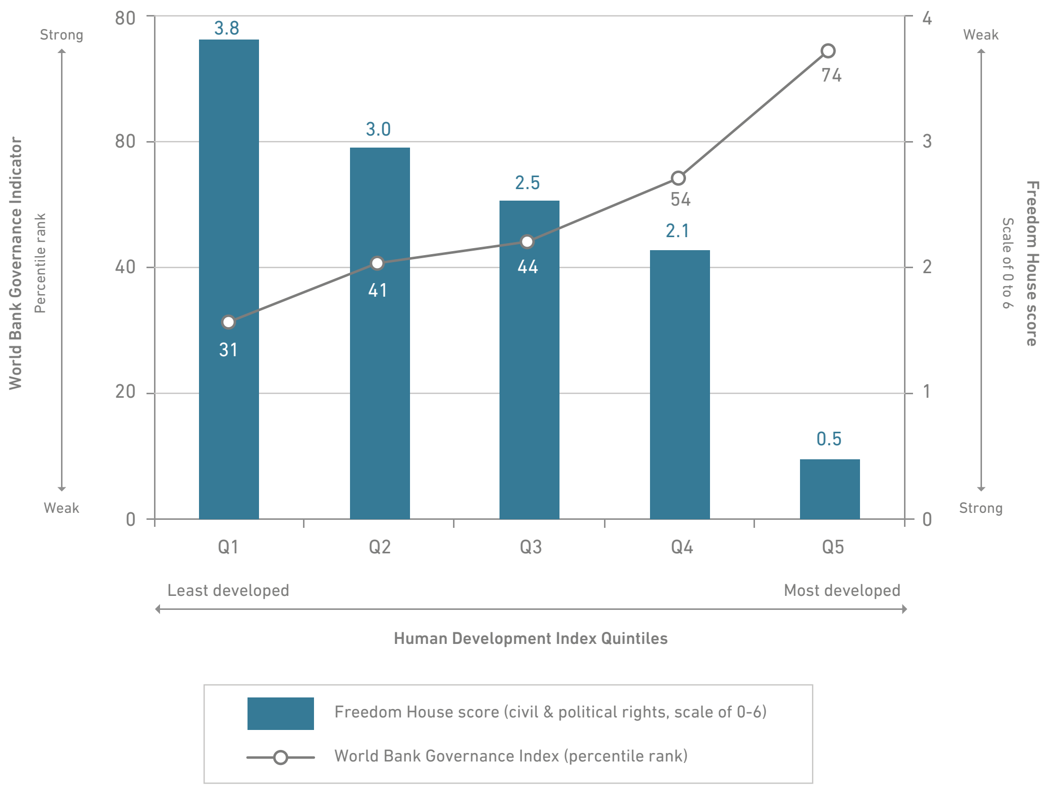


Exhibit 11. Countries that are less developed tend to have weaker policy environments and protections of citizens' rights.[31] Given all the concerns related to data and AI in industrialized countries, it is likely that the scope for abuse is higher in developing countries.

[31] Human Development Index quintile data from http://hdr.undp.org/en/data. World Bank Governance Indicator data from http://info.worldbank.org/governance/wgi/index.aspx#home. Freedom House data from https://freedomhouse.org/content/freedom-world-data-and-resources.

6. Conclusions

The above analysis of food security, health, energy access and education shows that across the 38 most critical interventions (Exhibit 12), 19 have limited dependence on sophisticated data analytics, 14 can benefit from conventional analytical tools which have been in commercial use for at least two decades, and only 5 require advanced AI.

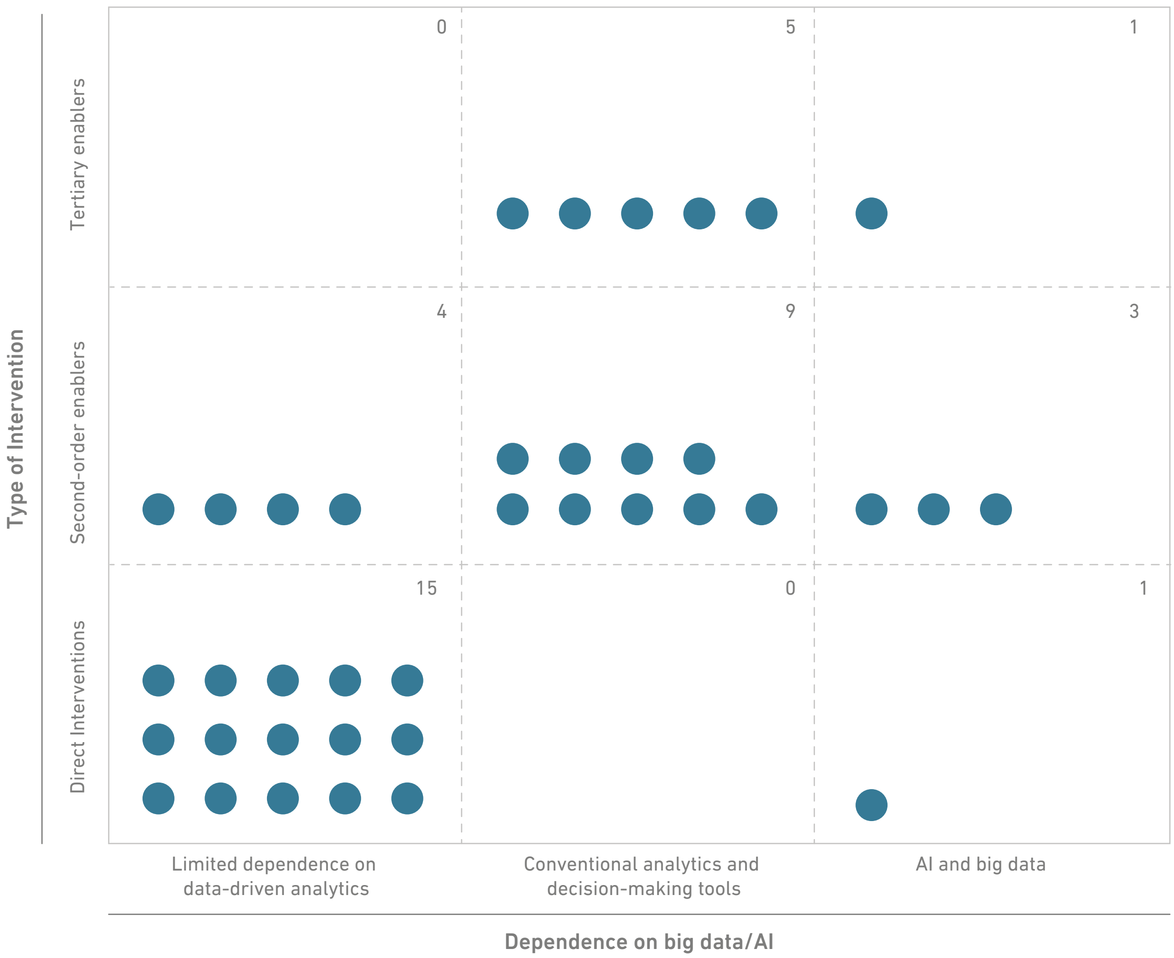


Exhibit 12. Of the 38 most important interventions required to address food security, healthcare, energy access and education in developing countries within the SDG timeline, 19 do not require much in the way of sophisticated data analytics; 14 interventions can be significantly improved using conventional analytics; and 5 require AI and big data.

There is certainly a lot of value in collecting, analyzing and utilizing reliable data, to glean insights, make better decisions, and even automate tasks when there is inadequate human capacity for them. However, in the context of human development—especially within the 2030 timeline for the SDGs—there is much more foundational work that needs to be done before sophisticated data and AI tools can be used for effect.

For example, while there is tremendous value in automated diagnosis from x-ray and ultrasound images, it will only have impact if there are enough affordable x-ray and ultrasound machines being deployed within reach of target populations. Similarly, while it will be very valuable to develop accurate risk models for crop insurance, the models will have limited impact without enough cash reserves to provide insurance in the first place, and without mechanisms to deliver and administer insurance policies. Therefore, it is important to fully understand the deployment context for such tools before investing heavily in them.

For stakeholders whose primary focus is on vertical SDG areas such as food security and health, data and AI interventions should only be considered as a part of integrated strategies for the vertical areas, so that there are clear linkages translating the improved information to better execution of direct interventions.

Even as foundational interventions are implemented, there is significant long-term value in improving the data infrastructures of developing economies, so that critical information can be captured across a broad range of public and private services, and citizens and consumers can be empowered to hold their providers to account.

To that end, India's Aadhar and India Stack initiatives, while still relatively early in their implementation, are beginning to show that even in a country as large and complex as India, transitioning away from paper-based administration will make it easier for governments to improve services to their citizens.

As things stand, institutions aiming to impact the SDGs have four options:

i. Focus on first-order interventions like the ones identified in the earlier sections;
ii. Invest in conventional analytics solutions to improve decision-making for the first-order interventions;
iii. Make long-term investments in building data infrastructures; or,
iv. Solutions based on new-generation AI.

In the least developed countries and communities, it is likely that first-order interventions are most appropriate. In emerging economies which have witnessed some development successes, and already have mechanisms to collect some relevant data, each of the first three options can be valuable; indeed, they need not be mutually exclusive and can be implemented in tandem. However, we believe the AI option is the least likely to lead to impact in the foreseeable future. Once strong foundations are laid, perhaps AI can truly be the game-changer many hope it will be.

Appendix

Data Density Index scores for 136 countries

Country	Data Density Index					
	Business	People	Government	Infrastructure	Sum	Rank
United Kingdom	24.8	24.3	24.6	20.3	93.9	1
Luxembourg	23.4	23.4	22.8	24.1	93.7	2
Sweden	23.8	23.8	24.3	20.3	92.1	3
Iceland	22.9	25.0	22.7	20.9	91.6	4
Netherlands	24.2	24.1	23.9	19.3	91.6	5
Singapore	23.3	23.2	24.1	20.8	91.3	6
Switzerland	23.7	23.5	23.3	20.6	91.1	7
Denmark	22.3	23.7	25.0	19.9	90.8	8
Norway	23.6	24.1	23.4	17.9	89.0	9
United States	23.9	23.7	24.0	17.1	88.6	10
Finland	23.2	22.9	24.1	18.3	88.4	11
South Korea	22.6	23.4	24.6	17.8	88.3	12
Japan	23.6	22.7	24.0	17.8	88.1	13
Germany	22.9	22.5	23.9	18.7	88.0	14
France	21.8	22.5	24.0	19.3	87.5	15
New Zealand	22.7	23.2	24.1	17.3	87.2	16
United Arab Emirates	23.5	22.2	22.7	18.7	87.1	17
Australia	21.6	22.8	24.7	17.0	86.0	18
Estonia	23.6	23.1	23.2	15.9	85.8	19
Malta	20.1	22.3	21.9	21.2	85.5	20
Ireland	22.1	22.8	22.6	17.4	84.9	21
Belgium	22.3	22.3	22.1	18.0	84.8	22
Canada	22.9	22.5	22.6	16.7	84.6	23
Austria	22.4	22.0	22.7	17.5	84.6	24
Israel	22.9	22.8	21.9	16.8	84.5	25
Bahrain	20.2	22.3	22.2	17.9	82.5	26
Portugal	22.1	20.9	21.9	17.3	82.2	27
Spain	20.5	21.3	23.0	16.9	81.7	28
Lithuania	22.9	21.9	20.6	16.0	81.4	29
Qatar	23.3	21.9	19.5	15.9	80.5	30
Cyprus	18.6	21.9	21.1	16.7	78.4	31
Italy	18.5	21.0	22.4	16.3	78.2	32
Czech Republic	22.1	21.0	19.4	15.4	77.9	33
Slovenia	19.8	21.1	21.1	15.9	77.9	34
Latvia	21.5	21.5	19.1	15.5	77.6	35
Uruguay	18.6	20.6	21.5	16.0	76.6	36
Malaysia	23.4	20.4	19.6	12.9	76.3	37
Chile	20.9	20.2	20.1	13.9	75.2	38
Russia	19.0	20.4	21.8	13.9	75.0	39
Poland	18.6	19.3	21.7	14.8	74.4	40
Greece	17.1	20.0	21.4	15.9	74.4	41
Bulgaria	19.2	20.2	19.6	15.4	74.4	42
Kuwait	18.2	19.3	20.2	16.3	74.0	43
Slovakia	21.4	20.5	19.5	12.2	73.6	44

Country	Data Density Index					
	Business	People	Government	Infrastructure	Sum	Rank
Costa Rica	19.9	19.7	19.1	14.7	73.5	45
Hungary	19.7	19.7	19.8	14.0	73.2	46
Saudi Arabia	20.3	20.6	19.5	12.8	73.1	47
Kazakhstan	18.8	19.4	20.8	13.9	72.9	48
Croatia	18.2	20.2	19.2	14.6	72.2	49
Serbia	16.8	19.7	19.5	15.7	71.8	50
Brazil	18.9	19.5	20.0	12.9	71.3	51
Thailand	20.2	19.7	17.9	13.5	71.2	52
Montenegro	17.4	19.8	19.0	15.0	71.2	53
Romania	18.6	19.5	18.2	14.7	71.1	54
Azerbaijan	20.3	20.1	18.0	12.3	70.7	55
Argentina	16.0	20.2	20.0	14.1	70.4	56
Turkey	19.7	19.3	19.4	11.6	70.1	57
Georgia	17.1	19.2	18.8	13.9	69.1	58
Macedonia	18.7	19.9	17.2	12.4	68.2	59
South Africa	19.6	17.3	18.1	13.3	68.2	60
Trinidad and Tobago	16.9	19.6	17.6	13.9	68.0	61
Moldova	16.6	19.2	18.0	14.1	67.9	62
Panama	20.5	17.9	16.6	12.7	67.7	63
Mauritius	17.5	18.6	18.2	12.9	67.3	64
Oman	16.4	19.0	18.7	12.9	67.0	65
Colombia	19.1	17.5	18.8	11.5	66.9	66
Armenia	18.6	18.7	16.2	12.2	65.7	67
Mexico	18.8	17.2	18.6	10.4	65.0	68
Philippines	19.4	17.9	17.8	9.9	65.0	69
Ukraine	18.2	18.1	16.8	11.6	64.7	70
China	20.2	16.6	18.6	9.2	64.7	71
Mongolia	18.7	17.6	15.9	11.6	63.8	72
Albania	16.2	18.0	17.8	11.2	63.2	73
Jordan	19.6	19.1	15.2	9.3	63.2	74
Seychelles	16.6	17.3	16.8	12.1	62.8	75
Peru	18.1	16.2	17.7	10.8	62.7	76
Lebanon	15.3	19.4	15.1	12.1	62.0	77
Vietnam	19.4	16.3	16.2	10.0	61.8	78
Ecuador	18.0	15.7	16.7	10.5	60.9	79
Dominican Republic	18.6	16.4	15.6	9.9	60.6	80
Tunisia	15.6	17.1	17.1	10.5	60.4	81
Sri Lanka	19.9	15.8	15.7	8.2	59.6	82
Indonesia	20.5	17.2	14.4	7.3	59.3	83
El Salvador	17.7	15.6	14.9	11.1	59.3	84
Bosnia and Herzegovina	15.7	17.2	14.5	11.7	59.1	85
Cabo Verde	17.7	17.1	13.6	9.3	57.7	86
Guatemala	19.8	15.1	13.6	9.0	57.5	87
Jamaica	18.0	17.1	12.8	9.5	57.4	88

Country	Data Density Index					
	Business	People	Government	Infrastructure	Sum	Rank
Morocco	17.5	16.9	14.2	8.8	57.4	89
Iran	16.1	15.3	16.6	9.2	57.3	90
Venezuela	14.6	17.8	14.4	9.8	56.7	92
Namibia	18.1	15.6	12.4	9.8	56.0	93
Egypt	17.4	17.2	13.3	7.9	55.8	94
Kyrgyzstan	15.7	15.3	15.9	8.0	54.9	95
Honduras	19.6	15.1	12.2	7.7	54.6	96
Kenya	20.1	14.7	12.4	7.3	54.5	97
Botswana	16.3	16.1	11.6	9.8	53.8	98
Paraguay	15.3	15.3	14.4	8.7	53.7	99
Ghana	17.0	14.5	14.7	7.0	53.2	100
Algeria	14.2	15.3	11.5	10.6	51.6	101
Bolivia	14.7	14.0	14.5	8.5	51.6	102
Gabon	14.1	14.7	11.8	10.4	51.1	103
Cambodia	17.6	14.7	10.3	8.5	51.1	104
Guyana	16.8	14.6	11.8	7.0	50.1	105
India	16.9	12.3	15.5	5.4	50.0	106
Rwanda	18.9	12.7	12.5	5.6	49.8	107
Tajikistan	16.1	16.7	11.5	5.2	49.5	108
Nicaragua	15.0	13.0	11.6	8.6	48.1	109
Senegal	18.8	13.4	9.5	6.3	48.0	110
Bhutan	15.0	14.8	11.7	5.8	47.3	111
Bangladesh	15.4	12.4	13.3	5.6	46.7	112
Zambia	17.4	12.9	11.2	4.7	46.3	113
Nepal	14.7	13.2	13.0	5.1	46.0	114
Nigeria	17.3	13.7	10.4	4.4	45.8	115
Côte d'Ivoire	17.4	13.4	7.6	6.3	44.7	116
Gambia	16.1	13.1	8.1	7.2	44.5	117
Myanmar	13.3	13.2	9.1	8.5	44.1	118
Uganda	16.6	12.1	11.1	4.2	44.0	119
Cameroon	16.9	12.2	10.9	3.4	43.5	121
Zimbabwe	14.9	13.2	10.1	5.1	43.3	122
Laos	16.8	13.0	8.3	4.8	42.9	123
Mozambique	15.9	12.2	8.7	5.2	42.0	124
Pakistan	15.7	11.4	9.7	4.7	41.6	125
Liberia	14.9	15.0	7.5	4.1	41.5	126
Tanzania	15.2	10.5	10.7	4.6	41.0	127
Benin	16.2	11.5	8.9	3.5	40.1	128
Lesotho	13.9	11.1	8.1	5.7	38.8	129
Ethiopia	14.4	10.5	9.5	3.7	38.1	130
Mali	15.6	11.3	6.6	3.9	37.4	131
Mauritania	15.1	12.1	6.3	3.3	36.8	133
Madagascar	16.5	11.4	7.6	0.7	36.3	134
Malawi	14.0	10.6	7.4	3.3	35.2	135

Country	Data Density Index					
	Business	People	Government	Infrastructure	Sum	Rank
Haiti	13.3	10.9	8.3	1.6	34.1	137
Guinea	13.2	10.0	6.4	3.4	33.0	138
Burundi	11.3	8.0	8.2	3.9	31.3	141
Chad	10.9	7.7	3.4	1.0	23.0	144

INSTITUTE FOR
TRANSFORMATIVE
TECHNOLOGIES

The next generation of breakthrough technologies
to combat global poverty

www.transformativetechnologies.org