International Journal of Communication 14(2020), 4385–4405 1932–8036/20200005

AI4D: Artificial Intelligence for Development

SUPREET MANN[1]
MARTIN HILBERT
University of California, Davis, USA

We derive a conceptual bridge between technical concepts from the deep learning literature and natural metaphors for international development. We start with a rather technical review of four of the characteristic traits of deep learning technologies: representation, reuse, robustness, regularization (4Rs of deep learning). Based on the empirical evidence of 24 case studies, we derive four characteristics of the use of AI4D that align with the four technological traits, namely development foci on local and distance intelligence, and mirrored and detailed reality representations. In isolation, each one of the identified issues presents a plethora of opportunities to contribute to international development, especially to the attainment of the Sustainable Development Goals. However, in combination, they create a clear tension between a looming threat of a hegemonic intelligence indoctrination pushed by global economies of scale and the potential promise to not only honor but also celebrate local diversity with the help of flexible AI designs.

Keywords: artificial intelligence, machine learning, sustainable development, international development, deep learning

In line with a long research tradition focused on the use of information and communication technology for development (ICT4D), we explore the role of artificial intelligence (AI4D). Information and communication technology (ICT) has long been recognized as an important tool for international development. Decades of thoughtful academic research, often grouped under the shortcut "ICT4D" (Heeks, 2006, 2017; Unwin, 2009), has fueled multiple generations of national and international policy agendas ("eLAC Action Plans," n.d.; "World Summit on the Information Society," n.d.). This active line of research grew out of work often grouped under the term *development communication*, which focused on the diffusion of mass media in developing countries and goes back to work from some of the founders of today's communication discipline in the social sciences (e.g., Schramm, 1979).

Supreet Mann: sumann@ucdavis.edu
Martin Hilbert: hilbert@ucdavis.edu
Date submitted: 2019–06–05

[1] We thank Amy Hoang, Shangzhen Li, Julia Morris, Jenna Rutledge, and Sruthi Sivanandan of the Computational Communication Research Lab (C^2) at the University of California, Davis, for their research input for the elaboration of the case studies in the appendix.

As the digital paradigm evolves, so must the focus of this research. The most prominent recent change has been a shift in focus, from the proliferation of communication (1970s–1980s) and information (1990s–2000s), to the extraction of knowledge from the resulting data. Every two to three years, humankind deals with more technologically mediated information than it had since the beginning of human history (Hilbert, 2015). The only way to deal with this information deluge is by fighting fire with fire: using digital machines to make sense of the information provided by digital machines. Therefore, we have started to outsource the important task of interpreting and filtering digital information to intelligent algorithms.

In this study, we develop a framework that allows us to reason about the role of artificial intelligence (AI) for international development by considering the interplay between modern AI concepts and their potential applications. We start by discussing the characteristics of modern AI and then match these characteristics to opportunities and dangers of AI for international development.

Our theoretical review of modern AI focuses specifically on deep learning techniques. We distill four main concepts, referred to as the 4 Rs of deep learning: representation, reuse, robustness, and regularization. Reviewing these concepts leads to natural metaphors for the potential role that AI can play to tackle pressing development challenges. We derive this conceptual bridge by an empirical review of 24 case studies that illuminate how AI currently contributes to the fulfillment of nine of the United Nation's 17 Sustainable Development Goals (SDGs). We crystalize four general characteristics of the application of AI4D: (1) local intelligence, (2) distance intelligence, (3) mirrored reality, and (4) detailed reality. Each one matches conceptually with one of our 4 Rs of deep learning.

Combined, we obtain a conceptual framework for AI4D that allows looking at international development from the perspective of the technical particularities of the deep learning paradigm. All the identified characteristics have obvious positive effects on development outcomes, but on second thought, their combination could also lead to some fundamental threats to developing countries. We finish our analysis with a reflection on such possible threats to global inequality.

AI: The Theory

We start with the historical context of AI and then advance to its general architecture, all with the goal of identifying some of its main characteristics.

Current State of Affairs

Private sector companies agree that the global market for smart machines was around US$15–20 billion in 2018 and that its contribution to the global economy is soon to reach several trillion (Bughin, Seong, Manyika, Chui, & Joshi, 2018; International Data Corporation [IDC], 2018).

The leading role of AI in today's economy is pushed by dazzling advancements. Deep neural nets have managed to reduce the word-error rate in speech recognition from 26% to 4% between 2012 and 2016 (Lee, 2016), making them much better than human transcribers (Xiong et al., 2016). Deep convolutional neural networks identify skin cancer with accuracy that matches that of trained experts

(Esteva et al., 2017). When using this technology in a developing community that may not otherwise have access to healthcare, we see AI applications in real-life scenarios with real outcomes. The positive contribution of AI is not limited to one realm; we see its impact in medical advancements, human relationships, and an overall potential for social good.

The electric grid is in the hands of AI (Ramchurn, Vytelingum, Rogers, & Jennings, 2012); three of four transactions on the U.S. stock market are executed by automated trading algorithms (Hendershott, Jones, & Menkveld, 2011); and one in three marriages in America begins online (Cacioppo, Cacioppo, Gonzaga, Ogburn, & VanderWeele, 2013), making intelligent algorithms an undeniable player in sexual mating, energy, and financial transactions. And this widespread impact is unlikely to slow down as ICT rapidly advances because of the proliferation and expansion of AI to fit developing communities and their unique needs.

A Short History of AI

Although all this progress seems to have happened in a historical blink of an eye, intelligent machines have occupied human thought for 2,000 to 3,000 years, from depictions of robotic creations in the Talmud and Homer's *Iliad*. Most researchers place the birth of modern AI to the 1950s, related to Turing's famous formulation of the "Turing test" (whether a human can distinguish between human and machine behavior), and the so-called Dartmouth workshop from 1956, an eight-week-long brainstorming session that informed many of the general directions in the field during the subsequent decades. Participants of this workshop, such as the AI pioneer Herbert Simon, predicted that "machines will be capable, within twenty years, of doing any work a man can do" (Schreuder, 2014, p. 419). Another attendant, the AI pioneer Marvin Minsky, agreed, writing, "Within a generation . . . the problem of creating 'artificial intelligence' will substantially be solved" (Schreuder, 2014, p. 419).

By 1985, the global market for AI had reached more than $1 billion. During the 1990s and 2000s, the world of technological progress focused on the proliferation of information diffusion solutions by means of Internet connections, databases, and phones. This resulted in an information overload, and researchers started to look for computational solutions to make sense of the data deluge. The current breakthrough in AI dates back to a result from 2012, when Geoffrey Hinton and collaborators surprised the academic world by showing the power of so-called deep convolutional neural networks (in this case, for image classification; Allen, 2015). These are not based on expert systems fed with identified patterns (knowledge, grammar, decision rules, etc.), but on machine learning algorithms that discover patterns.

Today's AI: Machine Learning

Traditional AI systems, called expert systems, focused on automating insights gained by humans. To recognize a car, one would teach the machine the rules that define a car (four wheels, certain size, etc.). In contrast, modern AI systems adopted a learning approach more akin to how children learn: by examples, not by rules. A child learns to distinguish cars from motorcycles not by evaluating a series of rules, but by seeing different examples of each. This aims at identifying new patterns in data, not to match patterns against a given decision rule. The ability of AI to build its own knowledge is known as machine learning (ML), and it allows computers to make decisions that seem to be both situational and subjective. The resulting classification criteria are more flexible and natural than predefined rules (Halevy, Norvig, & Pereira, 2009).

Today, this kind of ML has almost become equivalent with the term *artificial intelligence*. Machine translation is an epitome of this trajectory of AI. Since the 1950s, digital heavyweights such as IBM, MIT, DARPA, and others all worked on encoding the rules of grammar and vocabulary translation into expert systems, much like an automated textbook of translation among different natural languages. The results could, at best, be used to support, but not substitute, human experts. In 2006, Google Translate launched a statistical ML translation engine. Google Translate does not apply grammatical rules like an expert system would, but is fed with a bilingual text corpus of more than 150–200 million words, and two monolingual corpora each of more than a billion words (Och, 2005). The machine learns the relationships itself. The result is that Google Translate now supports more than 100 languages at various levels and serves more than 500 million people daily.

Future Outlook on AI

Advancements of AI lead us to understand there are alternative understandings of intelligence distinct from human intelligence. It should not be surprising that we are discovering that the way evolutionary pressures designed human intelligence is just one of many possible implementations of a much larger and broader concept. Machines are currently discovering alternative ways to be intelligent as they are finding ways to promote learning outside traditional boundaries of human insight. This is the main driver behind the increasing complementarity between human and AI.

Deep Learning Architectures

Having established that modern AI is basically equivalent to ML, we now review some of the theoretical characteristics at AI's most important implementation, so-called deep learning, or deep neural networks (DNNs; Goodfellow, Bengio, & Courville, 2016). We aim to identify technological characteristics that lend themselves to tackling development challenges. We focus on four concepts articulated herein as the 4 Rs of deep learning: representation, reuse, robustness, and regularization. We will then relate these four AI-related technological characteristics to our framework of AI on development dynamics.

Deep Layers: Representation

One of the primary ways that AI machines are able to understand the situational and subjective nature of data is through representational learning. Representation learning is a set of methods that allows a machine to be fed raw input and to discover, from this input, representations that are needed for classification (LeCun, Bengio, & Hinton, 2015). Deep-learning methods are essentially representation-learning methods with multiple levels of representation that gradually result in representation at increasingly abstract levels.

Traditional ML algorithms are fed with certain features that represent some raw data. For example, a doctor interprets a scan image and feeds the observed features into the machine (the machine receives a representation of the image, not the image), which then makes suggestions for action (e.g., calculating the probability of requiring surgery). This requires doctors with specialization in medical imaging, which is costly and can be subjective. One solution is to use ML to discover not only the mapping from representation to output, but also the representation itself, which is called representation learning. Deep learning furthermore implies that the levels of features are learned from data and are not explicitly designed by human engineers. In other words, the machine is not only learning the data structure (traditional ML), but also part of its own high-level architecture.

Representation learning depends on particular factors of variation that help to separate each unique factor of the representation (Goodfellow et al., 2016). One of the central problems with this approach is that in many circumstances, some of the factors of variation influence multiple pieces of data, making it necessary to separate the factors of variation and ignore the ones that are insignificant. Deep learning solves the problem of separating the factors of variation by "introducing representations that are expressed in terms of other, simpler representations" (Goodfellow et al., 2016, p. 5). For example, Figure 1 shows the example of face recognition, as has been done millions of times in social networks such as Facebook and Instagram.

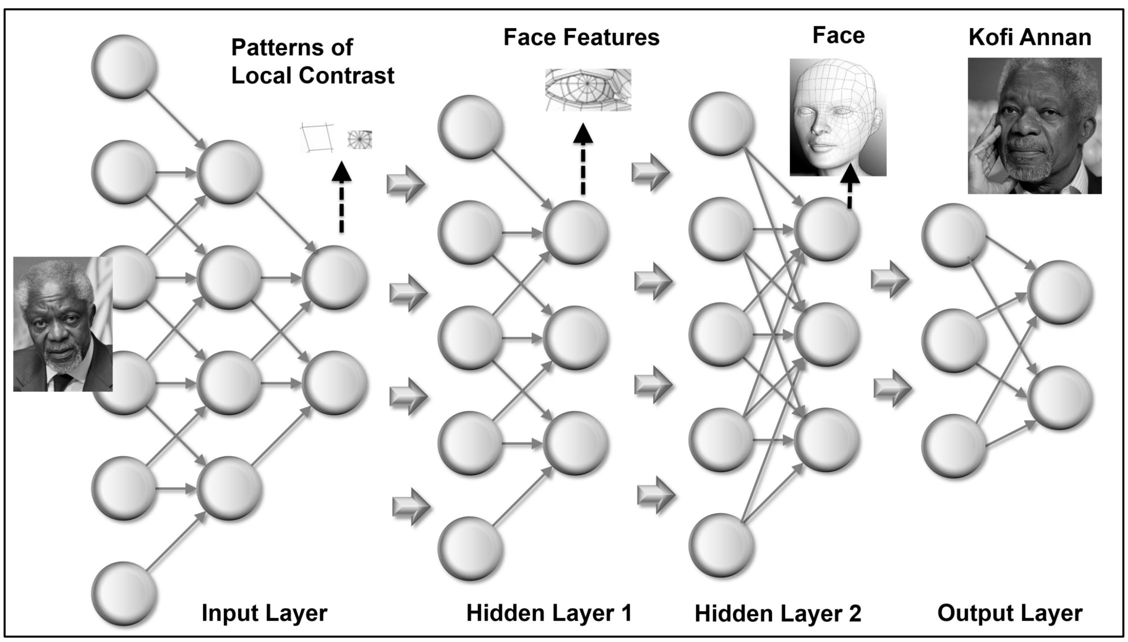

Figure 1. Schematic representation of face recognition through deep neural networks (images: Wikipedia.commons).

Thus, a deep-learning architecture is essentially a multilayer stack of simple modules, which are subject to learning (LeCun et al., 2015). The classic example of a deep learning model is the feedforward deep network, or multilayer perceptron (Goodfellow et al., 2016). A multilayer perceptron is a function that maps some set of input values to output values using a series of hidden layers that extract abstract features from the input or visible layer (Goodfellow et al., 2016). As shown in Figure 1, different layers learn different aspects of the whole, introducing flexible and robust modularity. To move from one layer to the next, a set of units calculate a weighted sum of their inputs from the previous layer and pass the result through a nonlinear function (LeCun et al., 2015). Backpropagation is often used to help achieve a goodness-of-fit optimum by giving a network the ability to form and modify its own interconnections.

Multitask- and Transfer Learning: Reuse

The important result of the multilayer, modular representation of knowledge is that it allows for better generalizations, as a "scheme for minimizing the generalization error of the prediction functions and deducing the biases with respect to the provided training set" (Yu, Zhuang, He, & Shi, 2015, p. 313). One can also focus on layer-by-layer training and then use the insights gained from one layer to improve tasks in another layer. The result is essentially a transfer of knowledge, whereas the modular nature allows for context-dependent adjustments without the need to start from scratch.

The vast collection of methods referring to multitask learning shares those parts of the model across tasks that capture a common pool of structure. The underlying assumption is that among the factors that explain the variations observed in the data associated with different tasks, some are shared across different contexts. For example, image recognition DNNs share learned features about lines, eyes, and faces on lower levels (see Figure 1). Online recommender systems learn to transfer shopper preferences among books, music, and consumer electronics. "The notion of re-use . . . is . . . at the heart of the theoretical advantages behind deep learning, i.e., constructing multiple levels of representation or learning a hierarchy of features" (Bengio, Courville, & Vincent, 2013, p. 1802). When this idea is implemented in a semisupervised setting, it is often referred to as multitask learning (Goodfellow et al., 2016), whereas it goes under the name of transfer learning when implemented through supervised learning (Goodfellow et al., 2016).[2]

This technique is extremely beneficial if there are significantly more data in one setting than in another, which seems useful when considering the inequalities typical for international development. The classical case in the literature is to train computer vision with images of house cats and then use the extracted features to detect wild and rarely appearing snow leopards (Yosinski, Clune, Bengio, & Lipson, 2014). It can even be used to approximate unprecedented scenarios for which no label examples are available (so-called zero-shot learning; Goodfellow et al., 2016).

Convolutional Neural Networks: Robustness

Convolutional neural networks (CNNs) are one type of deep, feedforward network that is considered easy to train and generalize and is one of the most common implementations of deep neural networks (LeCun et al., 2015). CNNs are designed to process data that come in the form of multiple arrays, such as

---

[2] ML is said to learn supervised, if the desired output is already known. One trains the machine to convert certain input into certain output through trial-and-error supervision. With unsupervised ML, the machine is given a certain theoretical framework and is asked to pick up patterns. Traditional principal component analysis is among the oldest unsupervised learning algorithms.

a color image made up of a 2D grid containing various pixel intensities. They have been tremendously successful in practical applications.

Convolutional nets are the greatest triumph of biologically inspired AI. They are based on the insight that some neurons respond to very specific patterns, and hardly to others, while being very robust and detail invariant when doing what they do (Hubel & Wiesel, 1968). Convolutional nets implement this by systematically making use of parameter sharing that involves at least two types of layers: convolutional layers and pooling layers. Whereas the role of the convolutional layer is to detect aggregations of features from the previous layer, the role of the pooling layer is to merge similar features into one. This way, even if an input image has millions of pixels, we can detect small, meaningful features with kernels that occupy only tens of pixels, and share these parameters. This is done by sliding overlapping windows of shared representation over the grid structure.

The important result is that this particular form of parameter sharing in convolutional nets causes the layer to be equivariant to translation. This means that if the input changes, the output changes the same way. One obtains the same representation of some input, even if it occurs earlier or later, or if it occurs shifted to the one side or the other. For example, this allows it to detect if a face is in an image, without getting lost in the details of which direction it looks, what the background or context is, and so on (see Figure 1). For human development, the concept of equivariance ensures that different inputs are represented efficiently and can be detected even if they reappear in a highly context-dependent and volatile settings (e.g., development).

Overfitting: Regularization

The challenge of making learning both robust and flexible points to the main difficulty related to the output of ML: deciding when to stop learning. The algorithm might learn details that are particular to the specific data set, but are not generalizable. In computer science lingo, this is known as the problem of overfitting, which stands for the idea that the algorithm learned more details than it should have learned. "Overfitting literally means 'Fitting the data more than is warranted'" (Abu-Mostafa, Magdon-Ismail, & Lin, 2012, p. 119). It happens automatically with learning patterns and is often subjective. Often the final purpose of the application defines which aspects are warranted and which are noise; this is often a subjective decision.

For the ML community, the most common way to deal with overfitting is known as regularization. "Regularization is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error" (Goodfellow et al., 2016, p. 117). *Regularization* is a broad term that includes many different methods, most of them being approximate heuristics. As a result, "regularization is as much an art as it is a science" (Abu-Mostafa et al., 2012, p. 126). It is an implementation of the similarly hand-wavy concept of Occam's razor, summarized by Einstein as "everything should be made as simple as possible, but not simpler" (O'Toole, 2011, para. 1).

AI4D: The Practice

Inspired by the achievements and architectures of modern AI systems, we now set forth to analyze how they relate to international development. We based our conceptualization of development on the United

Nations SDGs. We gathered 24 case studies of the use of AI for development. These case studies are presented in the online appendix and present a variety of viewpoints, with some case studies highlighting pilot projects or initial concepts and others showing actualized and tangible results. They address nine of the 17 different SDGs: Two address SDG.2 (zero hunger); 11 address SDG.3 (good health); three address SDG.4 (quality education); two address SDG.5 (gender equality); one addresses SDG.8 (economic growth); four address SDG.11 (sustainable cities); four address SDG.12 (responsible consumption and production); two address SDG.14 (life below water); and two address SDG.15 (life on land). Some of them simultaneously address several goals (see Table 1 and the online appendix).

The method of case collection was based on random-walk and snowballing principles (starting with one example, which leads to related ones) of written reports of adequate case studies. This makes the sample nonrepresentative of actual cases of AI4D, but without a global travel budget, we were forced to work with cases that had previously received attention and provided written insights. A cursory online search of AI technologies shows that these cases are generally well known. We do not claim that the selection of case studies is exhaustive or comprehensive; however, they did serve the function of helping us translate the technological characteristics of the prevailing ML paradigm into important aspects of international development.

While we analyzed the case studies according to the previously identified 4 Rs of deep learning, we realized that the collected case studies lined up with four general characteristics to frame the effects of AI on development dynamics (see Table 1). The first two refer to the location of information processing with AI, and the second two refer to the input and output of this processing with regard to empirical reality.

Transferring intelligence:

1. Local intelligence: AI systems can be autonomously applied locally, adjusting to local context and requirements.
2. Distance intelligence: Modern telecommunication networks allow highly trained AI systems to be applied at a distance.

Manipulating reality:

3. Mirrored reality: AI systems allow the creation of digital twins of aspects of reality, which enhance our understanding and allow us to replicate aspects of reality.
4. Detailed reality: The digital footprint provides ever more detailed maps of reality, and AI allows us to exploit the resulting details to foster development goals.

As shown in Table 1, the match between the 4 Rs and these four groups is not perfect, and sometimes subjective. By design, the different characteristics are not orthogonal to each other, but overlapping. For example, mirrored reality can be implemented locally or at a distance. In the same sense, robustness can be applied to representation or reuse.

Table 1. Characteristics of Deep Learning and AI4D, With 24 Case Studies) and Corresponding SDG.

	Representation	Reuse	Robustness	Regularization
Local	CS1 (SDG.2): Climate & crops analytics CS2 (SDG.3): Pharma research engines CS3 (SDG.4): Individualized education CS4 (SDG.4): Hidden learning patterns CS5 (SDG.5): Gender equality CS6 (SDG.11): Sustainable cities CS7 (SDG.11): Smarter cities CS8 (SDG.12): Risky pipe detection CS9 (SDG.12): Demand irrigation CS10 (SDG.15): Timber conservation	CS11 (SDG.3): Alleviating medical paperwork	CS1 (SDG.2): Climate & crops analytics	CS12 (SDG.8): Productivity enabler across sectors
Distance		CS13 (SDG.3): Malaria screening CS14 (SDG.2): Diagnostic support CS15 (SDG.2): Analyses of tuberculosis CS16 (SDG.2): Automated diagnosis CS9 (SDG.2): Cataracts detection CS17 (SDG.12): Real-time water supply		

Mirrored		CS18 (SDG.3&4): Informing and guiding pregnancies and girls' rights	CS19 (SDG.2&12): Plant-based replication of animal-based foods CS20 (SDG.11): Road repair CS21 (SDG.14): Ocean ecosystems CS22 (SDG.15): 3D model of Earth	
Detailed	CS14 (SDG.3): Chemical compound research	CS3 (SDG.4): Individualized education		CS20 (SDG.11): Road repair CS21 (SDG.14): Protecting endangered species CS23 (SDG.3): Predicting cardiovascular disease CS24 (SDG.3): Avoiding unnecessary surgeries

Note. See online appendix.

While overlapping, a possible 3D representation would be redundant, given the natural one-to-one lineup of the 4Rs with the four AI4D concepts. Our goal was to identify a general outlook to frame the AI4D discussion, not to elaborate an exclusive, objective, and exhaustive classification scheme. The framework aims at relating technological traits with development characteristics.

AI Representation for Local Intelligence

One characteristic that we frequently found in case studies is the possibility to implement intelligence locally, using representation learning, which automatically embraces local conditions and necessities. This is very promising given that one of the most frequent critiques of international development work is the (often subliminal) "one size fits all" or "best practice" mentality (Tödtling & Trippl, 2005). Representation learning allows for the ad hoc training of autonomous agents that consider the particularities of local conditions in remote areas. The digital footprint provides constant input in the form of a steady pipeline that fuels new discoveries by exploiting regional variances, particularities, and dynamics. This promises to enable innovative mechanisms sourced from local conditions. Having an automated way to learn about local particularities (i.e., representation learning) allows local actors to identify ever more tailor-made solutions for unique local conditions, augmenting economic and social efficiency domestically. This can facilitate unique solutions for local challenges.

Climate and crop analytics. An emblematic case is the local use of a ML algorithm (adopted from neuroscience) for the analysis of weather and local rice crop data in Colombia to analyze the effects of climate change (CS1 [SDG.2], see Table 1). The results were highly localized and provided recommendations on the level of different towns, highlighting the adoptability of this technology within a local context. The foresight helped 170 farmers in Córdoba avoid direct economic losses of an estimated $3.6 million and potentially improve productivity of rice by 1–3 tons per hectare.

Pharmaceutical research. Another case that shows the potential to learn tailor-made representations of knowledge that fits specific local contexts is Benevolent.ai, an AI that distills insights from the vast collection of pharmaceutical research. The vast majority of the increasing numbers of scientific papers published each day are unread and unknown to most scientists. After searching vast databases for a neurodegenerative disease, the AI recently suggested using compounds that researchers had never considered. Two of them unexpectedly worked better than the best available treatment drug at the time (CS2 [SDG.3]). There is the possibility that some pressing health problems faced by developing countries already have solutions, but without such AI, the remedy remains lost in the scientific information overload. This social good can only be made possible with technologies that are able to manage the vast amounts of existing information to create actionable responses for these specific communities that may not otherwise have the resources to do so.

Education. In the realm of education, AI solutions allow us to automate education and tutoring systems, allowing for low-cost solutions at scale. Highly structured subjects especially, such as language learning, software programming, or quantitative analytical skills, can be automated, including grading and performance tracking. Learning AI systems allows for the massification of an individualized education experience for structured course work (CS3 and CS4 [SDG.4]), which can then be applied in a specialized or unique local context.

Gender equality. Other ML technologies are being used to promote gender equality in the local workplace and the classroom via local intelligence. Doberman.io has employed ML and speech recognition to create an app that helps promote gender equality in the meeting room (CS5 [SDG.5]) by recording and analyzing speech during a meeting and then providing a visualization of speaker contribution by gender as the meeting progresses. The app aims to heighten awareness of the gender equality issue. Using this technology in communities where women have traditionally been marginalized may help to provide in-the-moment awareness regarding gender equality.

City sustainability. Much has been written about smart cities (in Latin America and beyond) and the application of cutting-edge AI to tackle urban challenges related to traffic, safety, and sustainability, which certainly falls into the category of localized intelligence (CS6 & CS7 [SDG.11]). HiBot employs an AI system designed to algorithmically locate where pipes are at risk of failure and employs inspection of pipes that have already been replaced by evaluating soil dynamics and electromagnetic forces coming from power lines (CS8 [SDG.12]). In the U.S., it has already detected hundreds of thousands of water pipe bursts per year across the country. This allows for conservation of water through prevention of leaky pipes. The application applies the technology to local infrastructure conditions, providing specified outputs based on community needs.

Water/forest conservation. Naturally, efforts in environmental protection can also benefit from adjusting AI to local conditions. Ecological conditions are often too unique and complex to transfer models one-to-one. For example, AI was used to govern drip irrigation system work by inserting a network of ground sensors into the soil to discover plants' irrigation needs, monitor demand, and optimize water use remotely (CS9 [SDG.12]). Neural nets were then used to effectively learn optimal irrigation schedules for the local conditions. The state government of Rio de Janeiro used ML on documents, databases, and satellite imagery to learn that more than 40% of the forest management operations in a certain area likely involved severe breaches of the law between 2007 and 2015 (CS10 [SDG.15]), showing just how adaptable these technologies can be for local communities.

AI Reuse for Distance Intelligence

Despite all tailor-made attention, which has certainly been neglected in many development projects in the past, it is also true that development dynamics contain a considerable common pool of shared factors. The arising synergies provide a fertile ground for the application of different kinds of multitask- and transfer learning methods. What we call distance intelligence is the ability for AI technologies to supplant resources in fields that were previously understaffed or underresearched, with the help of telecommunicated intelligence.

Health/diagnoses. One of the pioneering applications of distance intelligence is the use of AI in the health sector, such as for automated distance education and distance diagnoses to treat a number of ailments, including congenital cataracts, tuberculosis, and breast cancer. This can be used to provide access to medical intelligence in remote and underserved regions, nationally or internationally.

A total of 300–500 million cases of malaria illness occur annually, of which 1.1–2.7 million are fatal. In developing countries, the lack of access to accurate diagnosis is largely due to a shortage of expertise, coupled with a shortage of equipment. Findings of a recent survey carried out in Uganda show that only half of rural health centers have microscopes, and of that half, only 17% have staff with the training necessary to use the microscopes for malaria diagnosis (CS13 [SDG.3]).

Image processing and computer vision techniques have been used to identify parasites in blood smear images captured through a standard microscope. Algorithms used for other imaging purposes, such as face detection, can be transferred to recognize the malaria plasmodia in blood smear images captured using mobile phones attached to portable microscopes (CS13 [SDG.3]).

Google's DeepMind Health project can be used to interpret test results and learn which types of treatments are most effective for different patients (CS14 [SDG.2]). Although DeepMind was founded in London with the immediate goal of streamlining the United Kingdom's national health system, the technology has the potential to influence communities worldwide as it seeks to support an existing healthcare system and become a self-sustaining initiative.

The use of AI to diagnose and process medical images does not have to be fully automated; it is also often used to complement the work done by medical professionals, helping to save time and eliminate costly misdiagnoses. Zebra Medical Vision (Zebra-Med) has created a service called Zebra AI1 that uses algorithms to examine medical scans for just $1 a scan (CS14 [SDG.2]). The results are then passed on to radiologists, saving them time in making a diagnosis or requesting further tests.

AI Robustness for Mirrored Reality

The ability for AI to quickly learn new representations in different contexts in a robust way is perhaps one of the greatest promises of modern AI (Goodfellow et al., 2016). Many of the recent advances in AI allow for it to be used in a variety of new settings. The idea of robustness suggests that a shared representation of features can be examined from either the convolutional layer or the pooling layer. Because the convolutional layer detects aggregates of features and the pooling layer merges similar features, robustness allows one to examine similar or overlapping features either from above or below. The effect is that new cases do not exactly have to be identical to previous ones in order to be understood by the AI.

What we call mirrored intelligence refers to a group of applications that replicate aspects of reality and thereby augment the information related to these aspects. This allows the machine to create additional meaning by cross-referencing related aspects and creating yet unseen scenarios on these aspects of reality. The ability to apply various applications requires that AI be both robust and adaptable in its managing of views and perceptions.

Road repair. Autonomous cars are one of the most visible implementations. AI uses 3D maps to help vehicles make real-time decisions. By mapping scenarios on to the existing field, autonomous cars are able to reason among multiple options to determine the best course of action. Such applications require the robust and flexible processing of concepts, under equivariant translation. In an effort to improve safety, the same ideas can be applied to roads rather than cars, with information gathered (sometimes by drones, other times by car-mounted cell phone cameras) about road construction and maintenance (CS20 [SDG.11]). AI can be used not only to identify problems, but also to simulate new scenarios based on empirical driving behavior, leading to the development of new standards and infrastructure designs (CS20 [SDG.11]).

Plant-based food replication. Going one step further, beyond mirrored and virtual realities, AI is also being used to replicate the design of real-world atoms- and molecule-objects. Food replication is being used to combat global hunger. NotCo developed an AI program called Giuseppe, which uses and replicates the molecular composition of animal-based foods to determine which vegetables would create a food with similar taste, texture, and even smell (CS19 [SDG2&12]). Not Mayo is made mostly from basil, peas, potatoes, and canola oil, instead of eggs and canola oil, but is said to taste and have a texture almost exactly like normal mayonnaise. The resulting products are economically cheaper and much less taxing on the environment than animal farming.

Gender equality/girls' rights. In many practical applications, the idea of mirroring reality is often implemented as a mix of local and distance intelligence. This shows that the categories in our classification system from Table 1 are not exclusive. An entire group of game-based simulations falls into this category. We studied two representative cases of this category (CS18 [SDG3&4]). For example, the Half the Sky Movement develops mobile phone simulation games to raise awareness of the general audience regarding issues that women and girls are facing. In nine minutes, women and girls play out the adventure of managing a healthy and successful pregnancy, compressing the nine-month process into a short game experience, all guided by interactions with preprogrammed intelligent machines.

AI Regularization for Detailed Reality

Perhaps one of the most fundamental ways that AI is able to provide detailed insight on specific areas of development is by fine-graining our understanding of reality through analyzing unprecedented big data sources in a new way with greater nuance. Advancing into a more detailed representation of reality increases the risk of overfitting. The machine might learn particular aspects of a circumstantial situation that was unique and will never reoccur exactly in this way. Regularization allows the intelligent system to continue learning within specific contexts without overfitting to a specific situation. This allows a system to be more efficient while still accounting for the warranted details of the system input.

Road repair/protecting endangered species. Collecting fine-grained detail has proved fruitful in a number of ways that all address development goals. By monitoring infrastructure, AI contributes to road safety and leads to new road designs that can influence the way drivers behave (CS20 [SDG.11]). If these insights are sufficiently regularized to be generalizable, it allows us, for example, to collapse ecosystem models in order to map dependence on subsistence fisheries, thereby fostering sustainable development (CS21 [SDG.14]).

Predicting disease/avoiding unnecessary medical procedures. It is important to emphasize that the detailed data are necessary, but not sufficient to reap the benefits. Modern ML techniques, like deep learning, are necessary. For example, using the same data, CS23 (SDG.3) presents a case in which neural nets correctly predicted 7.6% more patients who developed cardiovascular disease than other more traditional statistical techniques. Given that 31% of all deaths worldwide are attributed to cardiovascular disease (World Health Organization, 2017), this suggests that 2.4% of all global deaths can be better predicted thanks to neural nets (an estimated 1.4 million deaths per year). Such accuracy can also be used to reduce healthcare costs. Our final case, CS24 (SDG.3), shows a situation in which ML was able to reduce the number of unnecessary surgeries for breast cancer by more than 30% compared with existing approaches. It is this combination of detailed data and the power of ML that allows for generalizing insights into new cases in a reliable way.

Discussion: Development at the AI Crossroads

We started with a historical and theoretical review of modern AI and identified four characteristics, the 4Rs of deep learning. Studying 24 case studies, we found that they align with four characteristics of international development. In isolation, each of them provides ample evidence for bright opportunities to foster the development agenda with AI, especially the SDGs. However, in combination, it turns out that they lead to a tension between global efficiency and local needs. We discuss the arising challenges in this final section.

Global Efficiency and Local Diversity

In theory, the very nature of modern AI exemplifies the ideals of local context dependency and global coherence. Computer scientists meticulously balance representation learning of new patterns with multitask- and transfer learning of known patterns, and the regularization of noisy features with feature robustness despite varying details. The figurative balance between global synergies and local particularities is at the heart of the power of ML. Our application of these principles to the practice of international development not only provides an eloquent analogy, but also faces socioeconomic pressures in the form of cultural and political transaction costs, economies of scale, and social cohesion. The tension between global efficiency and local diversity is not necessarily contradictory, but establishes a tension that must be addressed to ensure that localized needs are not pushed aside for general productivity.

Most ML is done within the industrialized context, given that the process can be very costly. Naturally, the resulting intelligence learns the patterns of the data it was trained on. Despite all theoretical ambitions of eradicating the one-size-fits-all model with flexible AI, economies of scale create a strong pressure to adopt local conditions to the imported AI behavior for the sake of economic efficiency. The result is a digital indoctrination of "one-AI-fits-all."

This starts with conflicting foci of the most urgent AI solutions. Different countries face different health epidemics. Priorities of developed and developing countries are not the same. Additionally, different countries face different variations of the same epidemic. Genetic mutations adjust to the genetics of the host, so an AI trained on cancer cells from hospitals in New York and Berlin might lead to dangerously confusing diagnoses when applied to remote regions of the developing world. Something similar accounts for many solutions that aim at improving safety. Roads are built differently in different regions of the world, so an AI trained in interpreting drone images from some regions might reach dangerously unreliable conclusions when economic efficiency urges their application to developing contexts. Needless to say, self-driving cars trained in Shanghai and London would be utterly confused when entering many traffic situations of a developing metropolis in Africa or Southeast Asia.

It is therefore important that when examining how today's AI machines learn, we consider differences between different AI technologies and how these differences can influence their applicability in various situations. If a particular AI technology is more "humanlike" in its learning ability, maybe it is better suited for a particular application within the context of the 4Rs, such as representation, whereas a technology that is focused on multitasking may be more useful for reuse. While we have provided clear explanations of the 4Rs and highlighted how different case studies fit in these categorizations, parceling them apart in this way requires some subjectivity when deciding which of the 4Rs is most applicable in each case. The online appendix delves more deeply into the case studies and highlights the nuances of each case and its multifaceted qualities.

History has provided countless examples where economic and cultural hegemony has led to the extinction of local values, culture, habits, and development goals. Although modern AI provides the theoretical possibility to celebrate diversity, such designs are not initially favored by economic incentives. It is cheaper to reuse a one-size-fits-all solution. The arising threat is a global indoctrination of intelligence from the producer to the consumer of AI.

Regulating International AI Regularization

The design of global AI systems that balances the trade-off between global efficiencies and local contexts comes down to finding the line between those results that are generalizable and those that are not. Taking an AI that works in one context and trying to apply it to other contexts without considering differences and limitations is a clear case of overfitting. For the ML community, "regularization is our first weapon to combat overfitting" (Abu-Mostafa et al., 2012, p. 126). The term *regularization* is adequate for our purposes because it carries the metaphoric double-meaning of the need for regulation of a process that ensures global diversity in a world where AI solutions take an increasing share of all decisions.

If economic principles favor the one-AI-fits-all model, economic incentives and social institutions would need to be designed to balance this pressure with context-dependent automation and the adoption of local needs. This presents an uphill battle against economic efficiency and therefore cannot purely follow market mechanisms. The ML community does something similar. Identifying shared parameters that fit many contexts is sometimes set as an a priori goal (e.g., Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012). This "regularizes each unit to be not merely a good feature but a feature that is good in many contexts" (Goodfellow et al., 2016, p. 260). Of course, a solution that works for all cases is often as useless as one that works for none.

Taking advantage of synergies while not neglecting local particularities is usually solved with a multilevel approach, and deep learning lends itself to it (see Figure 1). Different levels identify communalities and differences. Learning shared factors largely facilitates the adequate and noninvasive application of distance intelligence. For example, an AI that has learned the insight that "no one shall be subject to torture" (UN General Assembly, 1948) is already effectively restrained to guide decisions in questions related to labor- and women's rights, and educational systems.

It is therefore essential that one consider the potential negative impact that AI technologies might have on some of the SDGs, particularly those related to human rights. Raso, Hilligoss, Krishnamurthy, Bavitz, and Kim (2018) highlight many of the risks that AI might pose to human rights by stressing positive, negative, and indeterminate impacts that technologies in specific domains (such as criminal justice, education, and online content moderation) might have on particular human rights markers. For example, they highlight that introducing AI technologies in the criminal justice system has the potential to negatively impact an individual's right to a fair public hearing and right to be considered innocent until proven guilty. Similarly, AI technologies that impact online content moderation (such as enforcing standards or quality of publications) might have negative consequences for freedom of opinion/information (Raso et al., 2018).

In theory, AI could learn about common cultural, social, and political norms. It can be programmed to complement restricted search spaces for specific constituencies among different populations. Over the coming years, intelligent machines will inevitably learn the hierarchical architecture that constitutes the complex multidimensional preference structures of what we call global norms. The promise is that the result will present a coherent, moral, and ethical framework, with a certain level of shared beliefs and values, while at the same time emphasizing diversity and multiculturalism. The hierarchical layers of deep learning are a natural way to encapsulate the representation of both generic human values and specific human customs and preferences. Modern deep learning AI provides a tangible way to implement and celebrate this naturally existing hierarchy in socially embedded human preference structures.

While theoretically possible, the practical implementation of such nuanced hierarchical models faces the uphill battle against the brutal economic incentives offered by economies of scale of digital products. The silver lining is that, just like all other technologies before it, AI does not comply with technological determinism. It can be socially constructed (MacKenzie & Wajcman, 1999). This makes it especially important that we consider how to implement AI technologies. Policy considerations and regulatory frameworks govern how specific AI technologies might be implemented in differing contexts. For example, autonomous vehicles, which have the ability to severely cut down on automotive accidents, are governed by legal parameters, policy considerations, and even value systems as we are forced to consider what it means to be the "driver" and the nature of being "in control" of a vehicle (Acosta, 2018).

A number of challenges related to AI application in the public sector exist. Wirtz, Weyerer, and Geyer (2018) highlight four major dimensions of AI challenges: AI technology implementation, AI law and regulation, AI ethics, and AI society. These dimensions provide us with important considerations, especially as we consider how AI can be applied in local contexts with communities that need different approaches.

Global Negotiations and Local Efforts

In this arising negotiation between global efficiency and local diversity, the only controllable variable for developing countries is the level of proactivity of their role. The question is about the weight they will bring to the anecdotal negotiation table regarding the hegemony of AI. The combination of our theoretical and practical analyses results in the hope that local communities will begin investing in building their AI capacity to avoid being swamped with solutions that may be inadequate and, in the worst case, damaging for them. Furthermore, the involvement of local communities is essential to ensure the success of the AI technologies. When participants who executed ICT4D programs were not included in the design of these programs, the project suffered from major failures despite being community-based and well planned (Brown & Mickelson, 2019). To avoid a similar outcome with developing AI technologies, it is paramount that future research include community partners that would directly benefit from the technology in every step of the project.

On the one hand, this is facilitated by cloud services. Many aspects of AI solutions are openly available, including deep learning suites such as Google's TensorFlow and Facebook's PyTorch. Even if developing countries do not produce the technology, they could use it to produce local knowledge, which is what matters most for reaping the benefits of AI4D. The reality in developing countries places many self-inflicted roadblocks in the way of this opportunity. This includes institutional factors and data scarcity. As one example, a small startup called BlackBox Solutions in Guatemala pursues the vision to take advantage of openly available tools like Google's TensorFlow to adjust ML to local contexts. Unfortunately, its country's legislation prevents it from implementing remote work models to employ programmers and from fundraising a round of capital (Sandel, 2018). Even if those institutional hurdles are taken, their daily work often consists of trying to extract data from PDF documents handed to them from local municipalities.[3] It is ambitious to aspire to train globally competitive AI with such scarce and expensive data input. Ready-made imported solutions from developed countries might be more economically convenient. Although this study is an important qualitative first step in examining how AI can be used for global good, it does not place an emphasis on randomization and representation and is limited in its examination of AI use. We are looking forward to many more examples and case studies that in the future will shed more light on the nascent topic of AI4D.

These and many other policy tools will play a crucial role in shaping the identified tension, and therefore the aspirations, of AI4D. Although each case study presents significant potential for furthering the goals of AI4D, each comes with its own set of limitations that are not discussed in this examination. Going further into the arising policy options is certainly beyond the scope of this study. It will require much more detailed and comprehensive consideration. This study aimed to present a general framework to approach the incipient discussion of the nascent field of AI4D. The discussion itself will accompany us for decades to come.

References

Abu-Mostafa, Y. S., Magdon-Ismail, M., Lin, H. (2012). *Learning from data*. Pasadena, CA: AML Book.

Acosta, A. J. (2018). *3 practical tools to help regulators develop better laws and policies*. Retrieved from http://cyber.harvard.edu/sites/default/files/2018-07/2018-07_AVs04_1.pdf

Allen, K. (2015). How a Toronto professor's research revolutionized artificial intelligence. *The Star*. Retrieved from https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. doi:10.1109/TPAMI.2013.50

Brown, S., & Mickelson, A. (2019). Why some well-planned and community-based ICTD interventions fail. *Information Technologies & International Development*, *15*, 49–61.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Modeling the global economic impact of AI.* Retrieved from https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

Cacioppo, J. T., Cacioppo, S., Gonzaga, G. C., Ogburn, E. L., & VanderWeele, T. J. (2013). Marital satisfaction and break-ups differ across on-line and off-line meeting venues. *Proceedings of the National Academy of Sciences*, *110*(25), 10135–10140. doi:10.1073/pnas.1222447110

[3] Personal interview, MH with BlackBox cofounder Nery Fernando Guzman at Hotel Barcelo, Guatemala City, October 1, 2018.

eLAC Action Plans. (n.d.). In *Wikipedia*. Retrieved June 15, 2019, from https://en.wikipedia.org/w/index.php?title=ELAC_Action_Plans&oldid=844452921

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. doi:10.1038/nature21056

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8–12.

Heeks, R. (2006). Theorizing ICT4D research. *Information Technologies and International Development*, *3*(3), 1–4.

Heeks, R. (2017). *Information and communication technology for development (ICT4D)*. London, UK: Routledge. doi:10.4324/9781315652603

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, *66*(1), 1–33. doi:10.1111/j.1540-6261.2010.01624.x

Hilbert, M. (2015). *Quantifying the data deluge and the data drought* (SSRN Scholarly Paper No. ID 2984851). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2984851

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*. Retrieved from https://arxiv.org/abs/1207.0580

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*(1), 215–243. doi:10.1113/jphysiol.1968.sp008455

IDC. (2018). *Worldwide spending on cognitive and artificial intelligence systems will grow to $19.1 billion in 2018, according to new IDC spending guide*. Retrieved from https://www.idc.com/getdoc.jsp?containerId=IDC_P33198

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi:10.1038/nature14539

Lee, A. (2016, March 18). *Geoffrey Hinton, the "godfather" of deep learning, on AlphaGo*. Retrieved from http://www.macleans.ca/society/science/the-meaning-of-alphago-the-ai-program-that-beat-a-go-champ/

MacKenzie, D., & Wajcman, J. (1999). *The social shaping of technology*. Buckingham, UK: Open University Press.

Och, F. J. (2005, September). *Statistical machine translation: Foundations and recent advances*. Tutorial presented at the MT Summit 2005, Phuket, Thailand. Retrieved from http://www.mt-archive.info/MTS-2005-Och.pdf

O'Toole, G. (2011). *Everything should be made as simple as possible, but not simpler*. Retrieved from https://quoteinvestigator.com/2011/05/13/einstein-simple/

Ramchurn, S. D., Vytelingum, P., Rogers, A., & Jennings, N. R. (2012). Putting the "smarts" into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM*, *55*(4), 86. doi:10.1145/2133806.2133825

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. Y. (2018). Artificial intelligence & human rights: Opportunities & risks. *Berkman Klein Center Research Publication*, *6*. http://dx.doi.org/10.2139/ssrn.3259344

Sandel, A. (2018, October 2). Este emprendedor guatemalteco está reinventando la big data. [This Guatemalan entrepreneur is reinventing big data]. *Centroamerica in Forbes México*. Retrieved from https://www.forbes.com.mx/este-emprendedor-guatemalteco-esta-reinventando-la-big-data/

Schramm, W. (1979). *Mass media and national development 1979*. Retrieved from http://unesdoc.unesco.org/images/0003/000370/037073eb.pdf

Schreuder, D. A. (2014). *Vision and visual perception*. Bloomington, IN: Archway.

Tödtling, F., & Trippl, M. (2005). One size fits all? Towards a differentiated regional innovation policy approach. *Research Policy*, *34*(8), 1203–1219. doi:10.1016/j.respol.2005.01.018

UN General Assembly. (1948). *Universal declaration of human rights* (217 [III] A). Paris.

Unwin, T. (Ed.). (2009). *ICT4D: Information and communication technology for development*. Cambridge, UK: Cambridge University Press.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615. doi:10.1080/01900692.2018.1498103

World Health Organization. (2017). *Cardiovascular diseases*. Retrieved from http://www.who.int/cardiovascular_diseases/en/

World Summit on the Information Society. (n.d.). *Wikipedia*. Retrieved June 19, 2020, from https://en.wikipedia.org/w/index.php?title=World_Summit_on_the_Information_Society&oldid=774075739

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., . . . & Zweig, G. (2016). *Achieving human parity in conversational speech recognition*. ArXiv:1610.05256 [Cs]. Retrieved from http://arxiv.org/abs/1610.05256

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 27* (pp. 3320–3328). Red Hook, NY: Curran Associates. Retrieved from http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf

Yu, W., Zhuang, F., He, Q., & Shi, Z. (2015). Learning deep representations via extreme learning machines. *Neurocomputing*, *149*, 308–315. doi:10.1016/j.neucom.2014.03.077

Online Appendix

A 33-page appendix is part of this study and features details about the 24 case studies and their specific relationships to the United Nations Sustainable Development Goals. The online appendix is available at https://osf.io/kjge4/?view_only=bfd6e4f0be254d40a23f58f4518709a1.