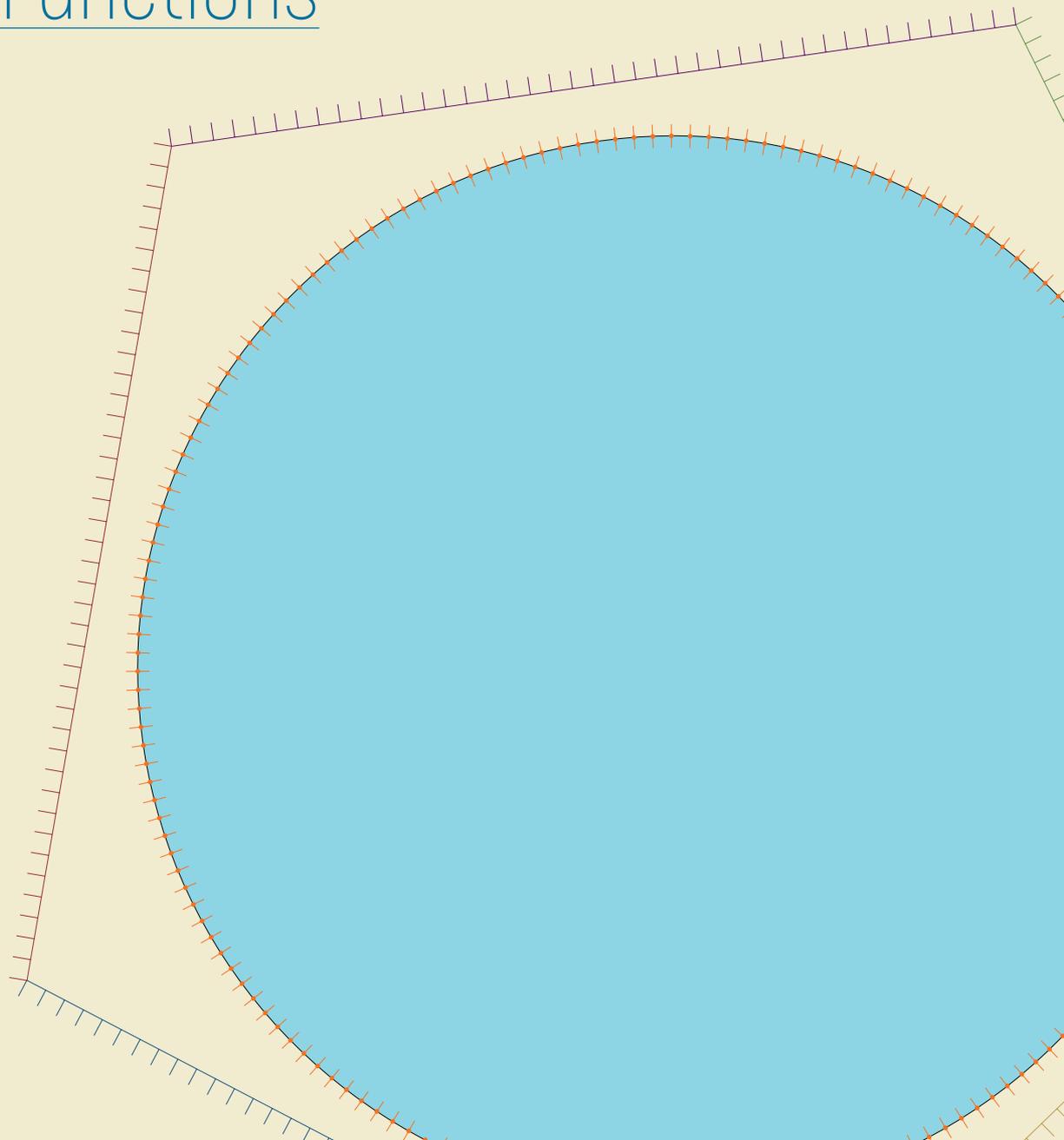


Institutional and
Technological
Design
Development
Through Use of
Case Based
Discussion

Arindrajit Basu,
Elonnai Hickok and
Amber Sinha

Regulatory
Interventions
For Emerging
Economies
Governing The
Use Of Artificial
Intelligence In
Public Functions



Introduction

Background and Scope

The use of artificial intelligence (AI) driven decision making in public functions has been touted around the world as a means of augmenting human capacities, removing bureaucratic fetters, and benefiting society. Yet, with concerns over bias, fairness, and a lack of algorithmic accountability, it is being increasingly recognized that algorithms have the potential to exacerbate entrenched structural inequality and threaten core constitutional values. While these concerns are applicable to both the private and public sector, this paper focuses on recommendations for public sector use, as standards of comparative constitutional law dictate that the state must abide by the full scope of fundamental rights articulated both in municipal and international law. For example, as per Article 13 of the Indian Constitution, whenever the government is exercising a “public function”, it is bound by the entire range of fundamental rights articulated in Part III of the Constitution.

However, the definition and scope of “public function” is yet to be clearly defined in any jurisdiction, and certainly has no uniformity across countries. This poses a unique challenge to the regulation of AI projects in emerging economies. Due to a lack of government capacity to implement these projects in their entirety, many private sector organizations are involved in functions which were traditionally identified in India as public functions, such as policing, education, and banking. The extent of their role in any public sector project poses a set of important regulatory questions: to what extent can the state delegate the implementation of AI in public functions to the private sector?; and to what extent and how can both state and private sector actors be held accountable in such cases?.

AI-driven solutions are never “one-size-fits-all” and exist in symbiosis with the socio-economic context in which they are devised and implemented. As such, it is difficult to create a single overarching regulatory framework for the development and use of AI in any country, especially in countries with diverse socio-economic demographics like India. Configuring the appropriate regulatory framework for AI correctly is important. Heavy-handed regulation or regulatory uncertainty might act as a disincentive for innovation due to compliance fatigue or fear of liability. Similarly, regulatory laxity or forbearance might result in the dilution of safeguards, resulting in a violation of constitutional rights and human dignity. Therefore, we have sought to conceptualize optimal regulatory interventions based on key constitutional values and human rights that the state should seek to protect when creating a regulatory framework for AI. To devise these interventions, we identify a decision-making framework consisting of a set of core questions that can be used to determine the extent of regulatory intervention required to protect these values and rights.

We have arrived at the framework by identifying key values and rights, and analyzing AI use cases to understand how different uses and configurations of AI can challenge these values and rights. Specifically, the paper examines:

1. Use of AI in predictive policing by law enforcement;
2. Use of AI in credit rating by means of establishment of the Public Credit Registry (PCR) in India; and
3. Use of AI in improving crop yields for farmers.

This paper is divided into three sections. In the first section, we look at various models of regulation. In

the second section, we expand on the use cases we chose to study in detail and the policy. In the third section, we identify core constitutional values that any regulatory framework on AI in the public sector should look to protect. In this section, we also highlight key regulatory interventions that need to be made to protect these values by developing a set of guiding questions.

We chose to work on the Indian ecosystem for three substantive reasons, apart from the convenience of geographic proximity, which allowed us to conduct our primary research. First, in terms of public policy advancement, we feel that working in India is important, as the technology and its governance frameworks are both in their nascent stages and the potential for the use of these technologies and their impact on the populace, especially those emerging technologically, is immense. Second, the constitutional framework in India on key issues such as privacy, discrimination, and exclusion has both a legacy of jurisprudence and is at a critical juncture, as they evolve and adapt with respect to emerging technologies. Finally, we believe that focusing on India allows us to make a unique contribution to the existing literature, as it charts out a potential regulatory model for other similarly placed emerging economies.

Our framework limits itself to decision making by a regulator when designing or deploying the AI solution. It does not delve into the adaptive regulatory strategy that needs to be devised as the AI project is implemented. It is also not an exhaustive framework, as many context-specific questions will alter its application. The objective is limited to framing broad questions that can guide specific regulatory interventions as decision makers choose to adopt AI.



Methodology

From the outset, we realize that the term “artificial intelligence” is used in multiple ways, and its definition is often contested. For the purposes of this paper, we define AI as a dynamic learning system where a certain level of decision-making power is being delegated to the machine (Basu & Hickok, 2018). In doing so, we distinguish AI from automation, where a machine is being made to perform a repetitive task.

The first stage of our research involved studying three applications of AI in public functions. Through primary interviews and desk research, we sought to understand:

- How the decision was arrived at to devise an AI-based solution
- Relevant policy or political enablers or detractors
- What preparatory research or field work was done before implementing the solution
- How the data was gathered and collected
- Impact assessment frameworks or evaluation metrics used to determine the success of the project by the developers and implementers
- External assessments of the impact

Using what was learnt from these case studies, we created a decision-making framework that relied on key threshold questions, as well as possible regulatory tools that could be applied.

Section I: Regulatory Models for AI

Privatizing Public Functions

Across the world, activities traditionally undertaken by the state, including running prisons, policing, solving disputes, and providing housing and health services, are increasingly being delegated to private actors, often either private firms operating transnationally (Palmer, 2008) or quasi-governmental actors (Scott, 2017a). It is not only a shift in the extent of legislative discretion but the creation of formal and rule-based arrangements that were not needed in the welfare state model, where the state delivered all services directly (Scott, 2017a). Braithwaite’s conception of a regulatory state combines state oversight with the commodification of service provision, where the citizen is treated as a consumer (Braithwaite, 2000). Businesses must deliver services with state oversight, but the extent of oversight and the modes of regulation must be determined contextually (Scott, 2017b).

The increasing privatization of public functions throws up two key constitutional questions. First, to what extent can public functions be delegated to a private actor? Little jurisprudence exists on this, as there have been very few challenges to privatization across jurisdictions. The Indian Supreme Court in *Nandini Sundar and Ors vs State of Chattisgarh* (2011), which banned the state designated private police organization, Salwa Judum, held that “modern constitutionalism posits that no wielder of power should be allowed to claim the right to perpetuate state’s violence... unchecked by law, and notions of



innate human dignity of every individual” (Sundar and Ors v State of Chattisgarh, 2011). The Court went on to criticize the state of Chattisgarh’s “policy of privatization” that was the cause of income disparity and non-allocation of adequate financial resources in the region, which in turn was responsible for the Maoist/Naxalite insurgency. However, there was no clarification on what services are “governmental” and cannot be delegated. The only clear carve out was the state’s monopoly on the use of violence, which could under no circumstances be delegated. Although some indication of where to draw the line comes from the following dictum of the Supreme Court in *Nandini Sundar*:

“Policies of rapid exploitation of resources by the private sector, without credible commitments to equitable distribution of benefits and costs, and environmental sustainability, are necessarily violative of principles that are “fundamental to governance”, and when such a violation occurs on a large scale, they necessarily also eviscerate the promise of equality before law, and equal protection of the laws, promised by Article 14, and the dignity of life assured by Article 21.”

The Israeli Supreme Court in *Academic Center of Law and Business vs Minister of Finance* (2006) had also invalidated a statute allowing for the privatization of prisons by reading its Basic Law. The judges in the majority opinion did not embark on an inquiry into whether private prisons worked better than those run by the government (*Academic Center of Law and Business v Minister of Finance*, 2006). Instead, there was an assumption made that privatization was illegal because private actors inherently harmed human rights more than public providers.¹ The Court argued that only the state itself had the right to deprive people of their liberty and dignity. The minority opinion countered this proposition by claiming that if the private sector was in fact able to maintain better prison conditions than the public sector, then privatizing prisons may actually further

human dignity instead of undermine it.² This is a valid concern for emerging economies as there are various circumstances, including AI deployment, where private actors can deliver services more efficiently than an overstretched state. However, given the implications for human rights and dignity, it is conceptually difficult to draw an objective line on delegation. The Court “assumed there is no constitutional impediment to privatization of a vast majority of services provided by the state”.³

In the US, no bar to privatization exists and the market for private actors providing prison services is booming (Pelaez, 2019). In fact, a US appellate body judge has stated that a prisoner only “had a legally protected interest in the conduct of his keeper, not in the keeper’s identity” (*Pischke v Litscher*, 1999). This lack of clarity on the definition, scope, and delegation of public functions means that when deciding the extent to which an AI use case can be delegated to a private actor, a number of other context-specific factors must be considered. These will be developed and discussed in Section III.

The second constitutional question hinges on the extent to which the state or a private actor can be held accountable for a violation of fundamental rights. The state action doctrine in the US formulates an apparently clear principle: constitutional rights apply to the state and not to private action (except in certain situations, such as *Habeas Corpus*).⁴ State action, simply put, includes all government action which includes acts by the executive, legislature, and judiciary at both the central and state levels (Jaggi, 2017). However, the doctrine has a clear “public function” exception. As per this exception, a private actor may be considered a state actor if it “performs the customary functions of government” (*Lloyd Corp Ltd v Tanner*, 1972) or if it performs a function that is “traditionally exclusively reserved to the state” (*Barrows v Jackson*, 1953). The Indian Constitution is similar in that Article 12 states:

1. Para. 18 (Procaccia)

2. Id. ¶¶ 2, 4

3. Id. ¶ 65 (Beinisch) However, Justice Jowell did note that policing, defence, treaty-making, prosecution, and dissolving Parliament may be core governmental powers. ¶¶ 29–30

4. First articulated in *The Civil Rights Cases* (1883)

“Definition in this part, unless the context otherwise requires, the State includes the Government and Parliament of India and the Government and the Legislature of each of the States and all local or other authorities within the territory of India or under the control of the Government of India.”

The question of whether private actors performing “public functions” comes under “other authorities” has come up before the Supreme Court. Questions have revolved around the status of the Board of Control for Cricket in India (BCCI). In *Zee Telefilms vs Union of India* (2005), the Supreme Court held that the BCCI is not discharging a public function, although it did not reject the public function test. The dissenting judges in *Zee Telefilms vs Union of India* (2005) recognized that with privatization and liberalization, as governmental functions are being delegated to private bodies, these private bodies must safeguard fundamental rights when discharging public functions. In 2015, the Supreme Court held that the BCCI is, in fact, performing a public function and therefore can be held accountable under Article 12 (Sethia, 2015). More recently, the Supreme Court held that a private university can be held accountable for violation of fundamental rights, as they are performing a public function or public duty by imparting education (Francis Coralie Mullin v UT of Delhi, 1981). Therefore, it is fair to say that Indian courts have adopted the public function exemption. Yet, given the lack of clarity on the definition of “public function”, a context-specific approach is needed when ensuring that appropriate accountability, grievance redressal mechanisms, and liability are imposed in such cases. One test we recommend for the purpose of classification is linking the public function back to recognize aspects of the “right to life” enshrined in Article 21 of the Indian Constitution. The Supreme Court has held that “the right to life includes the right to live with human dignity and all that goes along with it, namely, the bare necessities of life such as adequate nutrition, clothing and shelter, and facilities for reading, writing,

and expressing oneself in diverse forms, freely moving about, and mixing and commingling with fellow human beings” (Francis Coralie Mullin v UT of Delhi, 1981). While recognizing that the magnitude and scope of this right is contingent on economic development, the Court stressed that the basic necessities of life, and the right to carry on such functions, are essential for basic human autonomy. Therefore, any entity carrying out a function that has implications for any of the functions described could be treated as a “public function”, although this cannot operate as a hard and fast rule.

Challenges to Regulating AI

Regulation is often designed to avert, mitigate, or limit risks (Haines, 2017) to human health or safety, or more broadly, to the effective functioning of a society. However, the risks that AI pose are only just being discovered and will continue to be realized as a greater number of use cases are designed and implemented. Importantly, the risks posed by AI cannot be determined only by evaluating the technology at hand. A genuine assessment of risk must contextualize the technology within the socio-economic, cultural, and demographic space within which it is being applied. The same AI technology or solution used for a specific use case in the defense industry may pose very different risks when used in the educational sector.

Scherer charts out four problems with regulating AI development ex ante (Scherer, 2016): “discreteness”, which means that AI projects could be developed in the absence of large-scale institutional frameworks; “diffuseness”, which entails that AI projects could be devised by a number of diffuse actors in various parts of the world; “discreteness”, which means that projects will use discrete components and the final potential or risk of the AI system may not be apparent until the system finally comes together; and “opacity”, which means that the technologies underpinning the system may be opaque to most regulators (Scherer, 2016).

Given these challenges, several academics have advocated applying Ayres and Braithwaite's proposition of responsive regulation to AI development (Terry, 2019). Simply put, responsive regulation suggests that appropriate regulatory interventions should be determined based on the regulatory environment and the conduct of the regulated (Ayres & Braithwaite, 1992). The crux of the idea lies in a pyramid of enforcement measures with the most interventionist command and control regulations at the apex and less intrusive measures such as self-regulation making up the base (Ayres & Braithwaite, 1992). For all matters, Ayres and Braithwaite believe it is better to start at the bottom of the pyramid and escalate up the structure if the regulatory objectives are not being met. This way, the government signals a willingness to regulate more intrusively while averting the negative impacts of more interventionist regulation at the very outset (Ayres & Braithwaite, 1992).

However, when deploying AI in public functions, moving from a spectrum of leniency to intrusiveness in all instances is fraught with risks to core constitutional values and human rights. This holds particularly true when the project is in its design stage or just about to be implemented, and the impact is not entirely known. We therefore advocate for "smart regulation" – a notion of regulatory pluralism that fosters flexible and innovative regulatory frameworks by using multiple policy instruments, strategies, techniques, and opportunities to complement each other (Gunningham & Sinclair, 2017). Based on certain threshold questions that help identify risks posed by a specific use case to core values, we attempt to provide guidance as to what different instruments, strategies, techniques, and opportunities could mitigate these risks associated with AI development and use.

Modes of Regulation

Broadly speaking, "regulation" can be conceptualized as governing with a certain intention across a number of often-complex situations (Doekler,

2010) where competing interests are at stake (Kleinsteuber, n.d.). Traditionally, regulation has been determined by the sovereign, although market actors are increasingly determining their own regulatory frameworks, either through self-devised codes of conduct or in conjunction with sovereign entities. The decentralization of regulation away from a solely government-driven model is being spurred on by the fact that governments have incomplete information and expertise, and do not have the financial or human resources to devise, implement, and enforce regulation when emerging technologies propel rapid change and consequent uncertainty (Guihot, Matthew, & Suzor, 2017).

Primary (Government-driven) Regulation

Traditionally, governments have various tools at their disposal to implement legislation. This includes nodality, authority, funding, and organization (Hood & Margetts, 2008). Nodality refers to the government's pivotal role as a receiver and distributor of vast sources of information, which enable it to ensure implementation of the law by detecting breaches and subsequently passing sanctions (Hood & Margetts, 2008). Authority bestows the government with the power to enforce sanctions and "demand, forbid, guarantee, and adjudicate" in a manner that is respected by all stakeholders (Hood & Margetts, 2008). In governmental regulation, implementation is through force and punitive sanctions for non-compliance, with the regulated not necessarily having a clear say in the framing of the regulation (Doekler, 2010). The treasure chest refers to the variety of resources, both monetary and infrastructural, at the disposal of the government to carry out any task (Hood & Margetts, 2008). Organization is the bureaucratic structure which enables the government to actualize the three other unique elements.

However, all of these elements may not necessarily apply to the multifarious nature of tasks that need to be examined when regulating AI-driven

solutions, particularly in economies as diverse and heterogeneous as India. The challenges in keeping up with the rapid pace of technological evolution have been better understood by private companies such as Google and Microsoft, who have taken the lead both in bank-rolling and implementing a variety of AI-driven solutions (Basu & Hickok, 2018). They possess the requisite expertise and human resources to conceptualize and incorporate various tools of regulation into the governance of AI. Therefore, in the regulatory domain, these companies are driving the rules of the game by creating codes of conduct for themselves and their peers in industry.

Peer Regulation or Self-regulation

Jessop describes self-regulation as a system of bottom-up governance that allows private actors to limit the role of regulatory bodies by adopting a “reflexive self-organization of independent actors involved in complex relations of reciprocal interdependence, with such self-organization being based on continuing dialogue and resource sharing to develop mutually beneficial joint projects, and to manage the contradictions and dilemmas inevitably involved in such situations” (Jessop, 2003). In a self-regulatory ecosystem, actors conceptualize and voluntarily comply with their own set of codes, thereby serving as a form of informal regulation, with no punitive sanction for non-compliance (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020). Self-regulation can be one of two types. The first, more standardized form, describes situations where industry-wide organizations set rules, standards, and codes for all actors operating in that industry. The second, voluntarism, occurs when an individual firm chooses to regulate itself and create its own code of conduct without any coercion (Gunningham & Sinclair, 2017).

Attempts at self-regulation have already started in the governance of AI. A recent study (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020) by the Berkman-Klein Center at Harvard University (hereinafter the

“Berkman-Klein study”) identified eight sets of “ethical” AI principles put forward by a range of multi-national companies, including Microsoft, Google (Pichai, 2018), and IBM. Each of these sets of guidelines espouse a set of principles that defer but fail to explicitly incorporate standards of domestic or international law (Basu & Pranav, 2019). For example, to protect the Right to Equality, the Google AI principles merely seek to avoid “unjust impacts on people, particularly to those related to sensitive characteristics”, without referring explicitly to the various contours of and jurisprudence related to the Right to Equality across jurisdictions.

As identified by the European Commission High Level Expert Group, even after legal frameworks have been complied with, “ethical reflection can help us understand how the development, deployment, and use of AI systems may implicate fundamental rights and their underlying values, and can help provide more fine-grained guidance when seeking to identify what we should do rather than what we (currently) can do with technology” (European Commission, n.d.). However, Mittelstadt argues that ethical frameworks are prone to fail to regulate AI solutions because unlike other fields where ethics are used as regulatory interventions, AI lacks (1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms (Mittelstadt, 2019). Further, ethical guidelines devised by multi-national corporations often do not apply in the specific societal or legal contexts across jurisdictions (Arun, 2019).

Therefore, reliance on self-regulation through ethical AI guidelines may not be adequate to appropriately regulate the variety of ways in which AI may be deployed-in public functions and to genuinely protect core values and human rights.

Type of intervention	Enforceability	Rigidity	Creation	Applicability
Legislation	Highest. Binding law, along with clearly defined sanctions for non-compliance.	Highest. Clearly defined standards of municipal law with any ambiguity ideally being resolved by the judiciary.	Top-down. Devised by the legislator with optional consultation.	Lowest common denominator. Would only prevent directly identifiable harms resulting from AI. Would also require production of adequate evidence and causality.
Co-Regulation	Middle. Decentralized regulatory process may lead to a binding outcome.	Not unique. Could be clearly defined or vague depending on the outcome.	Participatory. With government, civil society and industry meaningfully engage in this process.	May have wide or narrow applicability to actors, situations, and individuals depending on the context.
Self-Regulation	Lowest. Enforceable at the organizational level but not binding. Reliance on “soft sanctions” with no clearly defined sanctions for non-compliance.	Lowest. Clearly articulated frameworks with greater ambiguity and more scope for manipulation.	Participatory. Devised through high-level consultations among industry and civil society but with an absence of government actors.	All AI that is ethical is necessarily legal. However, ethical frameworks have a broader applicability to harms that are outside the rigid confines of the law.

Table 1: Modes of regulation

Section II: Use Cases of AI in Public Functions

This chapter revolves around the governance of specific use cases that we studied concerning the use of AI in public functions in India. As the definition of “public function” remains unclear, we adopted a broad remit of use cases – from core governmental functions, which channel the state’s monopoly over the use of violence (as discussed in Nandini Sundar), to credit rating, which is seeing increased private sector involvement and does not easily fit into the notion of a core state function such as lawmaking or policing.

The policy ecosystem in India has sought to promote AI adoption with a number of policy instruments, underscoring the need to instrumentalize AI and create broad stroke frameworks and focus areas. These include the discussion paper for the National Strategy on Artificial Intelligence, published by India’s government think tank NITI AAYOG (Kumar, Shukla, Sharan, & Mahindru, 2018), as well as the Report of Task Force on Artificial Intelligence (Department for Promotion of Industry And Internal Trade, 2018) – a task force set up by the Ministry of Commerce. There are three main policy levers we can take away from the National Strategy. First, it suggests that the government should set up a multi-disciplinary committee to create a national data market place, so that organizations looking to derive data-driven insights can benefit from this data. Second, it proposes an “AI+X” approach that articulates the long-term policy vision for India. Instead of replacing existing processes in their entirety, decision making on AI should always look to identify a specific gap in an existing process (X) and add AI to augment efficiency. Third, it envisions the use of India as a garage bed for emerging economies, which we feel is a risky approach as it treats Indian citizens as guinea pigs without considering the potential impact on constitutional rights (Basu, 2019). Instead, India can set the tone for emerging economies by devising appropriate regulatory interventions that bring the best out of the technology without posing significant harms.

Without delving into the appropriate regulatory strategy for each use case, we explain each by looking at the following questions:

- How the decision was arrived at to devise an AI-based solution;
- Relevant policy or political enablers or detractors;
- What preparatory research or field work was done before implementing the solution;
- How the data was gathered and collected;
- Impact assessment frameworks or evaluation metrics used to determine the success of the project by the developers and implementers;
- External assessments of the impact;
- Extent of involvement of the private sector;
- Regulatory framework in the sector.

Predictive Policing in Government/Law Enforcement

Predictive policing is making great strides in various Indian states, including Delhi, Punjab, Uttar Pradesh, and Maharashtra. A brainchild of the Los Angeles Police Department, predictive policing is the use of analytical techniques such as machine learning to identify probable targets for intervention to prevent crime or to solve past crime through statistical predictions (Berg, 2014). Conventional approaches to predictive policing begin by using algorithms to analyze aggregated data sets to map locations where crimes are concentrated (hot spots). Police in Uttar Pradesh (Sharma S. , 2018) and Delhi (Das, 2017) have partnered with the Indian Space Research Organization (ISRO) in a memorandum of understanding (MoU) that allows ISRO’s Advanced Data Processing Research Institute to map, visualize, and compile reports about crime-related incidents.

There are also major developments on the facial recognition front. The Punjab Police, in association with Gurugram-based start-up Staqu, has begun

implementing the Punjab Artificial Intelligence System (PAIS), which uses digitized criminal records and automated facial recognition to retrieve information on the suspected criminal (Desai, 2019). Staqu has worked with police in a number of other states, including Uttar Pradesh, Uttarakhand, and Rajasthan (Ganguly, 2020).

It is important to acknowledge that bias existed in policing well before data-driven decision making came into the picture. Studies conducted in several states point to a disproportionately high representation of minorities and vulnerable communities in prisons (Common Cause, 2018). Muslims in particular have been impacted by this trend and have also reported the highest rates of contact with the police among any community (17%) (Common Cause, 2018). Courts have often found that incarceration has taken place based on false implications, which highlights flaws in the decision-making processes adopted by the police (Common Cause, 2018). This causes potentially flawed feedback loops, where increased police presence in certain areas is also leading to more crime being detected, in turn, leading to further police surveillance.⁵

The thinking behind devising and implementing predictive policing systems appears to be trust in the improved accuracy that data-driven decision making can provide. One official is reported as saying that “the key to [predictive policing] is the massive data on previous crimes and how best our people are able to analyze and correlate them with the present crimes” (Sharma, 2017).

A detailed analysis of Delhi Police’s predictive policing by Marda and Narayan, entitled Crime Mapping, Analysis, and Mapping Systems (CMAPS), is very useful in understanding how this data is collected (Marda & Narayan, 2020a). The source of the input data was through calls received by the Delhi Police

Dial 100 call center. Unfortunately, the input data at this level is often flawed. The call taker is expected to enter the details of the crime into the “PA 100 form”, which records information received from the caller into one of 130 pre-determined categories, or into “miscellaneous” if it is too difficult to slot them in cleanly. If more than one crime is reported, such as purse snatching and murder, only the more grievous crime is recorded. This is then escalated to the “Green Diary”, which is often at the mercy of the police officer recording the incident. Police officers commonly believe that complaints by women are usually false (Marda & Narayan, 2020b). Marda and Narayanan’s study confirms that gathering this information has been selective and subjective. Among police officers there is “a general apathy towards individuals living in slums and more forgiving outlooks with respect to individuals living in posh parts” (Marda & Narayan, 2020c).

The systems are shrouded in opacity, with CMAPS being out of the remit of the Right to Information Act, and appear to lack standard operating procedures or grievance redressal mechanisms. There is no legislation, policy, or guidelines that regulate and guide the operation of these systems, and no framework for evaluation. Reports indicate that there was no preparatory work or empirical research undertaken by the police to identify how concerns raised by multiple studies in other parts of the world where predictive systems have been deployed might play out in India. As Marda and Narayanan point out, the greater number of calls from poorer parts of Delhi might not be indicative of a higher crime rate than the relatively richer areas, it could simply be a cry of desperation from vulnerable communities who do not have access to other governance institutions (Khanikar, 2018). Given the current state of data curation practices, data-driven decision making might not provide a fair or accurate outcome.

5. Insights gained from primary interview

While there has been considerable political excitement about the use of AI and machine learning in law enforcement over the last few years (Basu & Hickok, 2018), there has also been parallel discourse advocating a need for caution about the use of such techniques. This cautionary note is even more pronounced in the use of machine learning by the state for public functioning, particularly where it leads to decision-making that impacts individual rights and entitlements. The intended use of AI by law enforcement in India to infer individual affect and attitude, offers a ripe opportunity to consider the opacity of such techniques. Even though the framers of the constitution deliberately kept the words “due process of law” out of the Indian Constitution, subsequent years of jurisprudence have adopted versions of the US constitutional law doctrines of “procedural due process” and “substantive due process” within the meaning of “procedure established by law” under Article 21. In criminal law, statutes that define offences and prescribe punishments are considered “substantive”, while others relate to matters of process are considered “procedural”. It is now accepted law that a procedural law which deprives “personal liberty” has to be “fair, just, and reasonable, not fanciful, oppressive, or arbitrary” (Maneka Gandhi v Union of India, 1978). During investigations, as per the criminal procedure code, law enforcement officers can take certain actions on the basis of “reasonable suspicion” and “reasonable grounds”.

In the life cycle of actions by law enforcement agencies and the courts, starting from the opening of an investigation, followed by arrest, trial, conviction, and sentencing, we see that as the individual gets subject to increasing incursions or sanctions by the state, it takes a higher standard of certainty about wrongdoing and a higher burden of proof. Actions taken by law enforcement agencies, such as surveillance or arrests based on the use of sentiment analysis would be subject to the standard of due process. However, there is no way to judicially examine the reasonableness of such an action if the process is not explainable.

The standard in the US law for search and seizure under the Fourth Amendment is also of “reasonable suspicion”, and we can look at US jurisprudence around this term for guidance. This standard was defined as requiring law enforcement agencies to “be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that [actions]” (Terry v Ohio, 1968). In the case of informant tips, US jurisprudence considers an informant’s veracity, reliability, and basis of knowledge as relevant factors (Illinois v Gates, 1983). The standard of “reasonable suspicion” under the Fourth Amendment protection is not met by all tips. For instance, anonymous tips need to be detailed, timely, and individualized (Alabama v White, 1990). The grounds of reasonable complaint and credible knowledge in Section 49 of the Code of Criminal Procedure in India speak to a similar expectation of reliability and basis of knowledge.⁶ It has also been clearly held that “reasonable suspicion” is not the same as the subjective satisfaction of a law enforcement officer (Partap Singh (Dr) v Director of Enforcement, Foreign Exchange Regulation Act, 1985), and clearly requires a good faith element on the part of the law enforcement agency (State of Punjab v Balbir Singh, 1994). In the case of a reliance upon an algorithm to substitute the role of tips, it is therefore necessary that the legal standards which can test the reliability and basis of an algorithmic technique, its suitability to the context, and the relevance of the dataset in use are evolved. However, where these techniques are opaque, as Marda and Narayanan have demonstrated, would severely limit the capacity of both law enforcement agencies to make informed decisions, as well as the ability of the judiciary to examine their use. When a law enforcement officer relies on tips to arrive at a good faith understanding, there is a clear way for a reviewing officer or a judge to evaluate the nexus between the available facts, good faith understanding, and the decisions taken – this is the basis of the review. The same is not possible in the case of an opaque algorithmic tool.

6. Section 49 (1) (a) of the Code of Criminal Procedure states as follows: “When police may arrest without warrant. (1) Any police officer may without an order from a Magistrate and without a warrant, arrest any person (a) who has been concerned in any cognisable offence, or against whom a reasonable complaint has been made, or credible information has been received, or a reasonable suspicion exists, of his having been so concerned.”

There are also significant issues with judicial and law enforcement application of due process laws in India. For instance, despite having laws on admissibility and strict legal standards on what evidence is admissible, these rules are often set aside.⁷ Even more alarming is the legal position on warrantless arrests, where the courts have held that police officers are not accountable for the discretion of arriving at the conclusion of reasonable suspicion while conducting a search on a suspect.⁸ The lack of these protections make it harder to hold police accountable for excessive or unlawful use of predictive policing methods. Laws such as the Unlawful Activities (Prevention) Act (UAPA) are notorious for placing wide and unaccountable discretionary powers in the hands of law enforcement agencies (Khaitan N., 2019). In the UAPA, for instance, the term “unlawful activities” includes “disclaiming” or “questioning” the territorial integrity of India, and causing “disaffection” against India. The egregiously broad wording of such provisions come close to not just criminalizing unlawful acts but also objectionable beliefs and thoughts. In this context, the derivation of likelihood of an individual to commit crime through an opaque and unreliable technique such as predictive policing posits key challenges for decision makers.

Credit Rating

AI is being harnessed by lenders to calculate credit scores and develop credit profiles. With the use of AI algorithms that draw from various data entries, such as an individual’s banking transactions, their past decisions, their spending and earning habits, familial history, and mobile data, firms can make fast credit decisions for typical and atypical applicants (ICICI Bank, 2020). For example, Loan Frame uses AI and machine learning to examine a borrower’s profile and evaluate their creditworthiness (Loan Frame, 2020). Similarly, start-ups such as Lending Kart (2020) and Capital Float (2020) use AI to assess the creditworthiness of micro, small, and medium enterprises (MSMEs) to help reduce the risk of defaulting. Kaleidofin is another start-up that has

attempted to solve the many challenges of financial inclusion in rural and semi-rural areas. They have used algorithms to analyze a variety of data and “recommend a single, seamless package of insurance and investment solutions” (Randazzo, 2013).

Companies and public sector banks assert that using AI has enabled them to bolster financial inclusion by including those who lack a formal credit history (Vishav, 2019, as cited in Singh & Prasad, 2020). Flaws in credit rating have existed across countries for some time (Smith, 2018), with the creditworthiness of an individual being contingent on local social and cultural notions of who “ought” to get loans, rather than simple number crunching (Kar, 2018a). Known as redlining, these practices have had deleterious financial and social impacts on minorities, particularly the African-American community in the US (Pearson, 2017; Corbett-Davies et al., 2017).

In a detailed exposition of what she terms the “moral economy of credit” in West Bengal, Kar demonstrates that bias on conceptions of “credit-worthiness” are entrenched among loan-givers across micro finance institutions (MFIs) (Kar, 2018a). She argues that “capacity was invoked as an ethical judgment [by the loan officer] of a borrower’s ability to repay a loan, and was understood not through a seemingly objective analysis of financial data but through repeated exchanges with the borrowers during the verification process” (Kar, 2018b). She identifies five categories of exclusion driven by loan officers at microfinance institutions: religion, caste, class, language barriers, and location. Discrimination is “inter-sectional” (Kar, 2018b). “A number of Muslim dominated neighborhoods in Kolkata are discriminated against both because of their religion and because they are non-Bengali – largely migrants from the central Indian states of Uttar Pradesh or Bihar” (Kar, 2018b). The lack of data on individuals operating on the margins of or outside the formal financial system, combined with these entrenched patterns of exclusion, has ignited enthusiasm for data-driven decision making in this field.

7. See Umesh Kumar vs State of AP (2013) 10 SCC 591 (“It is a settled legal proposition that even if a document is procured by improper or illegal means, there is no bar to its admissibility if it is relevant and its genuineness is proved. If the evidence is admissible, it does not matter how it has been obtained. However, as a matter of caution, the court in exercise of its discretion may disallow certain evidence in a criminal case if the strict rules of admissibility would operate unfairly against the accused.”)

8. Section 165 of the Code of Criminal Procedure

Machine learning algorithms are trained on curated datasets often referred to as “training data”. For the purposes of fintech lending, this could be datasets that contain information about people’s behavior online, spending patterns, living conditions, and geolocation, etc. As mentioned above, some fintech companies in India have publicly acknowledged that the number of data points is often around 20,000 (Nag, 2016). Machine learning-enabled credit scoring works by collecting, identifying, and analyzing data that can be used as proxies, as mentioned above, for the three key questions in any credit-scoring model: a) identity, b) ability to repay, and c) willingness to repay (Capon, 1982). With the advent of big data and greater digitization and datafication of information, new data sources such as telecom data, utilities data, retailers and wholesale data, and government data are available. Traditionally, credit-scoring algorithms consider set categories of data, such as an individual’s payment history, debt-to-credit ratio, length of credit history, new credit, and types of credit in use.

The Reserve Bank of India is in the process of establishing the Public Credit Registry (PCR) for India – a comprehensive database of verified and granular information that will create a “financial information infrastructure” for providing credit at a national level. Chugh and Raghavan (2019) identified five limitations in the functioning of the existing information infrastructure, which the PCR seeks to remedy. These

include a lack of comprehensive data, fragmented information, dependence on self-disclosure by borrowers, authenticity of the data, dated information, and inefficiencies due to multiple reporting (Chugh & Raghavan, 2019). Speaking about the registry, Dr. Viral Acharya, Deputy Governor, explained that “in an emerging economy like India, it is always felt that the smaller entrepreneurs, mostly operating under the informal economy do not get enough credit as they are informationally opaque to their lenders” (FinDev Gateway, 2019).

With the introduction of new forms of data, the richness of data may theoretically increase the predictive power of the algorithm (Ranger, 2018). However, narratives on greater accuracy presume both the suitability of input data towards the desired output, as well as faith that past attributes or activities that are used as training data do not lead to unintended outcomes (Joshi, 2020). There have been concerns that a combination of a vast variety of data points and the correlations recommended by machine learning processes will produce discriminatory outcomes that are not apparent and cannot be scrutinized in a court of law (Langenbucher, 2020). When a model relies on generalizations reflected in the data, the final result for the individual will be determined by shared data on the relative group that the system assigns to them, rather than the specific circumstances of the individual (Barocas & Selbst, 2016). Algorithmic

credit scores can remove bias only as much as the data that fuels them. Often, an assessment of the assigned group is also flawed. The development of “risk profiles” for individuals by the car insurance industry is a useful example (Kahn, 2020). Data might indicate that accidents are more likely to take place in inner city areas where the roads are narrower. Racial and ethnic minorities tend to reside more in these areas, which effectively means that the data indicates that racial and ethnic minorities, writ large, are more likely to get into accidents. Software engineers are responsible for constructing the mined datasets, defining the parameters and designing the decision trees. Therefore, as Citrone and Pasquale put it, “the biases and values of system developers and software programmers are embedded into each and every step of development” (Citron & Pasquale, 2014).

The roll out of algorithmic credit rating in India must be preceded by studies that map the possible disparate impacts of this practice and avoid some of the adverse impacts that have been experienced in other countries. Some companies have started taking individual steps to conduct grassroots level efforts (Kaleidofin, n.d.), but a larger industry-wide effort that is supported and endorsed by the government would be useful given India’s depth and diversity. The government also needs to ensure regulatory certainty, so that start-ups are cognizant of the legal ecosystem within which they are operating.

Credit rating in India is governed by the Credit Information Companies (Regulation) Act, 2005 and the regulations issued in 2006 (Government of India, 2006). The Credit Information Companies (Regulation) Act, 2005, defines credit information as any information relating to the amounts and nature of loans, nature of securities taken, guarantee furnished, or any other funding-based facility given by a credit institution that is used to determine the credit-worthiness of a borrower. Given the variety of data that can be analyzed using algorithms, the definition might need revisiting (Goudarzi, Hickok, & Sinha, 2018).

As per Regulation 9.5.5 of the Credit Information Companies Regulation, 2006, it is mandatory for a bank that has rejected a loan on the basis of a credit information company report to:

- (1) Send the borrower a written rejection notice within 30 days of the decision, along with (2) the specific reasons for rejection and (3) a copy of the credit information report, as well as (4) the details of any credit information company that constructed the report. If the decision has been rendered by crunching data through algorithms, the results must be human scrutable to the extent that a coherent explanation can be provided.

Improving Crop Yields for Farmers

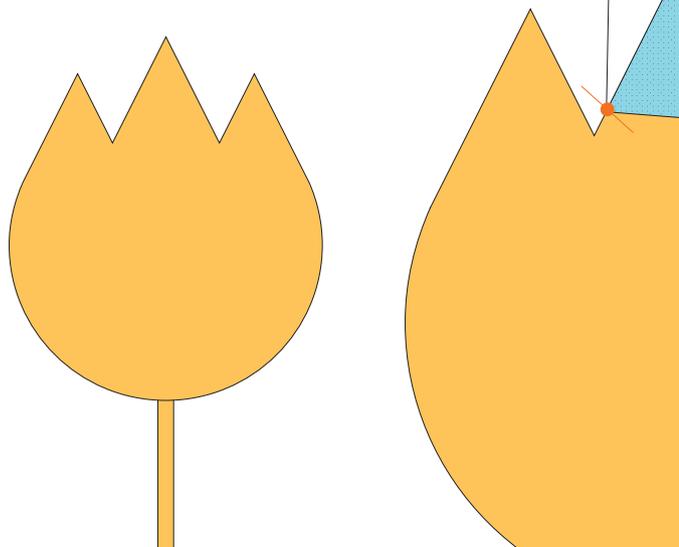
There has been a variety of initiatives taken by the government, in collaboration with the large technology companies, to equip farmers with more accurate information on weather patterns and ideal sowing dates for the generation of optimal crop yields (Gurumurthy & Bharthur, 2019).

IBM's Internet of things (IoT) platform has been used in many states in collaboration with NITI AAYOG – the Indian government's development think tank. The technology uses a "data fusion" approach which aggregates remote sensing meteorological data from The Weather Company, which is affiliated with IBM, along with satellite and field data (NASSCOM, 2018). In the state of Andhra Pradesh, Microsoft has collaborated with ICRISAT to develop an AI sowing app powered by the Microsoft Cortana Intelligence Suite. It sends advisories to farmers, providing them with information on the optimal date to sow by sending them text messages on their phones in their native languages. The government of Karnataka has signed a MoU with Microsoft to use predictive analytics for the forecasting of commodity pricing (UN ESCAP, 2019).

Despite being critical to India's economic development, the Indian agricultural sector continues to face a vast array of challenges (Indian Express, 2018): Some of them are associated with labor and resources, including migration to urban areas, overuse of groundwater, access to viable and quality seeds, a lack of balance in the use of fertilizers, and storage; infrastructure, including a lack of access to reliable credit, marketplaces, and technologies such as the Internet; and information, including a lack of access to reliable information about weather, markets,

and pricing (Nayak, 2015). Due to the information asymmetry in price modelling and forecasting, as well as weather and sowing conditions, specifically in Karnataka, the agricultural sector is characterized by a combination of drought-prone regions and areas that receive abundant irrigation (Deshpande, 2002). Compared to other states, Karnataka distinctively comprises a disproportionately large share of drought-prone areas (Deshpande, 2002). Farmer distress in Karnataka typically arises out of stress factors such as uncertainty in climatic factors and crop-prices (Deshpande, 2002). These conditions often have induced farmers to take miscalculated steps that result in onerous debts and sheer inability to meet family requirements (Deshpande, 2002). In addition, a study conducted in 2002 by the Karnataka State Agricultural Prices Commission identified that a large section of farmers (71%) did not end up selling their yield through regulated markets (Chatterjee & Kapur, 2016). This was because of an acute lack of knowledge (8%) of regulated markets (Chatterjee & Kapur, 2016).

Data-driven decision-making was targeted both by the state government of Andhra Pradesh and Telangana to address this specific gap (UN ESCAP, 2019). The implementation of the MoU was initiated through the development of an AI sowing app powered by the Microsoft Cortana Intelligence Suite, reported on June 9, 2016 (Reddy, 2016). Cortana Intelligence helps increase value in data by converting it into readily actionable forms (Heerdt, n.d.). This facilitates the expedient availability of information in achieving innovative outcomes within the agricultural industry. Using this intelligence, the app was able to interface



with models to forecast weather prepared by Where Inc. – a software company in the US. The app used extensive data mapping, including rainfall over the past 45 years in the Kurnool District (IANS, 2016; Reddy, 2016). The information was combined with data collected in the Andhra Pradesh Primary Sector Mission, popularly known as the Rythu Kosam Project (ICRISAT, n.d.). Launched with the objective of promoting productivity in the primary sector, the project involved the collection of household survey data relating, among other things, to crop yields (Charyulu, Shyam, Wani, & Raju, 2017). The combined data was downscaled in order to enable forecasting that could guide farmers in identifying the ideal week for the purpose of sowing (IANS, 2016).

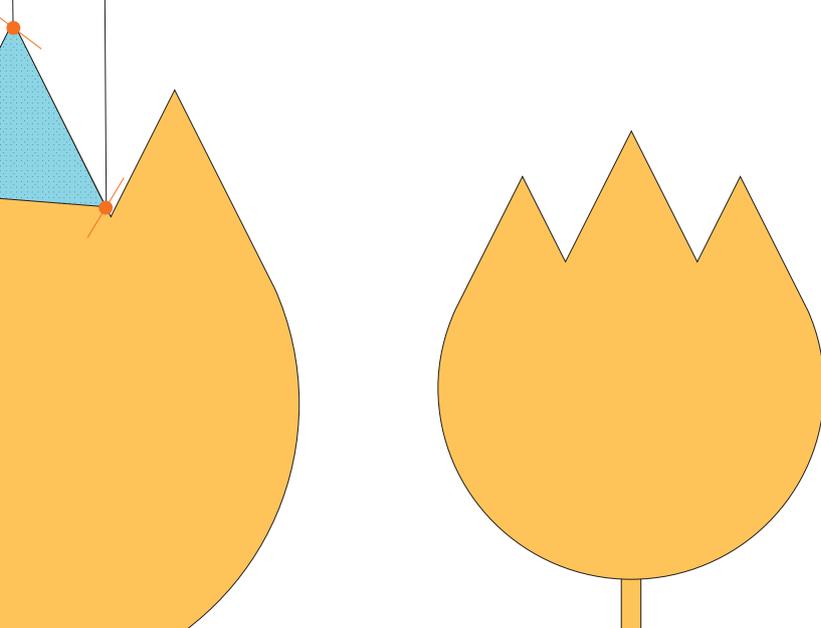
The datasets considered relevant for the AI solution include yield-related information, weather, sowing area, and production. Part of the data was manually collected from farms in 13 districts in Karnataka by field officers deployed by ICRISAT during the aforementioned Rythu Kosam Project. The information was made available to Microsoft's Azure Cloud (Express Web Desk, 2017) and subsequently downscaled to the village level in order to achieve the greatest possible precision, which was particularly useful for farmers in improving their decision-making capabilities. The machine learning software acquired by ICRISAT includes Cortana Intelligence and a personalized village advisory dashboard that uses business intelligence tools, both of which are prepared by Microsoft (ICRISAT, 2017).

In the pilot attempt implemented in Andhra Pradesh, the sowing period was estimated on the basis of datasets concerning the climate of the Devanakonda area in Andhra Pradesh, historically spanning a

period of 30 years (1986–2015) (ICRISAT, 2017). The estimation involved computing data to forecast a future moisture adequacy index (MAI) based on data concerning daily rainfall, which was accumulated and reported by the AP State Development Planning Society (ICRISAT, 2017).

However, there were infrastructure-related hurdles to the successful implementation of both projects. As of December 2017, the overall Internet penetration in India was around 64.84% (20.26% in rural areas) (Agarwal, 2018). This meant that the AI intervention had to be very targeted. Since 77% of the bottom quintile owned a mobile phone (Bhattacharya, 2016), the output needed to be sent as text messages and not through an app that required the user to have a smart phone.

The NITI AAYOG reported that both in Karnataka and Andhra Pradesh there was an increase in crop yield between 10–30% due to the ICRISAT sowing advisory app (NITI Aayog, 2018). As a result of the MoU, the government can reportedly get price forecasts for essential commodities three months in advance in order to decide the minimum support price (IANS, 2017). The first impact assessment conducted in Devanakonda Mandal in Andhra Pradesh reflected a significant increase (30%) per hectare for farmers using the app (ICRISAT, n.d.). However, there are no publicly available reports on a holistic impact assessment of this project. Furthermore, the calculations undertaken to arrive at the 10–30% increase have also not been furnished.



Section III: Regulatory Interventions

To determine the optimal levels of regulation, we have arrived at a set of principles that enable the policymaker to define how the solution can work in consonance with existing values and constitutional frameworks as applicable to emerging economies. Transformative constitutionalism is a new brand of scholarship in comparative constitutional law, which celebrates the crucial role of the state and the judiciary in bringing about emancipatory change and rooting out structural inequality. Originally conceptualized as a Global South (Christiansen, 2011) concept designed as a counter-model to the individual rights-driven model of Northern Constitutions, scholars have now identified emancipatory provisions in several Western constitutions, such as Germany (Hailbronner, 2017). India's Constitution is one such example. The origins of constitutional order in India were designed to "bring the alien and powerful machine like that of the state under the control of human will" (Khilnani, 2004) and to eliminate the inequality of "status, facilities, and opportunities" (Kannabiran, 2012).

Therefore, a transformational approach necessarily considers the power asymmetries between the decision maker, implementer, and affected party, respectively. The questions for guiding regulation are an entry point that remedy the inherent asymmetries which span out in a variety of contexts.

As public authorities begin to adopt AI into decision-making processes for public functions, and begin to determine the ideal form of intervention(s), the extent to and the way in which decision-making capabilities can and are delegated to AI need to be questioned from the perspective of its transformative impact on justice, civil liberties, and human rights.

A framework of high-level articulation of values and guiding questions can help to guide these determinations. We curated the values based on an assessment both of India's constitutional ethos and an evaluation of values and rights that might inherently be tested by and therefore need to be explicitly protected when there is algorithmic decision making. This section contains an explanation of how we selected these questions and how they protect these values. It then goes on to draw out what an illustrative regulatory strategy might look like in response to these questions.

Agency

Across jurisdictions, the concept of inherent dignity is connected to human agency – the capacity to make choices as one deems fit and pursue one's conception of a healthy life. Dignity reflected in agency does not require a specific set of criteria to define itself (Rao, 2013). It focuses on human capacities such as individuality, rationality, autonomy, and self-respect, and eschews focusing on the exercise of these traits (Rao, 2013). The Supreme Court of India has recognized the importance of the principle of autonomy in our constitutional schema and held that no discrimination by the state can undermine the personal autonomy of an individual (Bhatia, 2017).⁹ Of the instruments demarcating ethical uses of AI, 69% have adopted a principle of human control. This essentially requires that key decisions delegated to AI remain under human review with a "human-in-the-loop" (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020).

Where stakeholders have sufficient agency to inform their use or interaction with AI, there is a presumption of limited regulatory intervention required. The less the agency of a stakeholder in dealing with AI, the greater the regulatory intervention needed.

9. Naz Foundation vs NCT of Delhi, (2009) 160 DLT 277 (High Court of Delhi). ("The grounds that are not specified in Article 15 but are analogous to those specified therein will be those which have the potential to impair the personal autonomy of an individual... Section 377 IPC in its application to sexual acts of consenting adults in privacy discriminates a section of people solely on the ground of their sexual orientation which is analogous to prohibited ground of sex.", see Tarunabh Khaitan, 'Reading Swaraj into Article 15: A New Deal for the Minorities' (2009) 2 NUJSLR 419

Explanation

If adoption of an AI solution is mandatory, individual autonomy is immediately surrendered and the state determines the contours of individual agency. This is happening at present with the mandatory adoption of contact-tracing applications in light of the COVID-19 pandemic (Agrawal, 2020). During times of emergency or otherwise, if the state limits individual autonomy, then unique regulatory solutions that check the powers of the state must be deployed.

For AI solutions such as predictive policing, the primary users are state agents attempting to discharge their functions, whereas the impacted party is someone who is identified and evaluated by algorithmic decision-making. However, in the case of farmers receiving weather alerts, the farmer is both the primary user and the impacted party. To use another example, if the marketing and sales wing of a company uses sentiment analysis to analyze the user reviews of its products, the primary user, as well as the beneficiary or adversely impacted party of the analysis, is the company itself. On the other hand, if the same techniques are used for assessment of college application essays, the primary user is the university, but the parties who have to bear its adverse impact are the student applicants. Such a distinction must be made to determine if the potential risk of the algorithmic system is being borne by the stakeholders who choose to use it, or by other stakeholders who become unwitting victims of risks undertaken by others, and influences the impacted individual's ability to question the outcome or seek redress. Where parties choose to use systems marked by opacity and risk for commercial gains, there is a strong argument for regulatory restraint, unless the risks of such opaque decisions begin to percolate to others. In cases where the primary user and the impacted party are the same, there is a possibility for some opportunity for the user to play a role in deciding whether the inferences are used or not. In cases where they are not the same, the impacted party has no agency in this decision-making, and the further removed the role is, the potential for questioning this decision decreases when it is delegated to an algorithm.

Questions

The following questions can help guide determinations of agency:

- Is the adoption of the solution mandatory?
- Does the solution allow for end-user control?
- What is the relationship between the primary user and impacted party?

Recommended Regulatory Strategy

Adoption of the solution must be made mandatory only in exceptional circumstances. Compelling a farmer to adopt a technological solution constrains choice and undermines agency. Through primary regulation legislation or judicial decisions, we recommend that all states ensure that government entities at all levels adopt clear parameters for when any technological solution can be made mandatory. This must ensure that: (1) there is a pressing need in the public interest, (2) there is no reasonably available alternative, and (3) adequate measures of compensation, oversight, and grievance redressal are provided.

Even if the adoption of the solution is not mandatory, the power asymmetry between the user and impacted party needs to be closely considered. Where the power asymmetry is vast, such as police using AI to conduct surveillance in certain areas without the knowledge or consent of the people impacted, there needs to be far greater regulatory scrutiny. Ideally, this scrutiny should be multi-stakeholder and civil society groups, especially those representing vulnerable communities, and should be allowed to exercise vigilance by inputting into the design of the project before it is launched, auditing evaluation reports, engaging with targeted populations, and providing input as the project processes. Furthermore, training must be mandated for the public servants implementing the solution, thereby enabling them to understand the socio-economic complexities of those with whom they are engaging. Marda and Narayanan observed a lack of sensitization and empathy in the case of the Delhi police dealing with vulnerable communities (Marda & Narayan, 2020a) while Kar observed the same with loan officers passing judgement on "credit-

worthiness" (Kar, 2018a). Appropriate grievance redressal mechanisms that provide access for the vulnerable must be created. This should all be mandated through a top-down policy that is devised by the central government and made applicable to all government entities thinking of adopting AI solutions that have a great disparity between the end user and impacted party.

Equality, Dignity, and Non-discrimination

Background and Explanation

Human dignity is a core value recognized the world over, which the state should guarantee. In the Indian Constitution, dignity is mentioned in the Preamble and nowhere else. However, the Supreme Court has used the inclusion of the concept in the Preamble to interpret the guarantee of life and personal liberty to include a variety of traits associated with dignity. These include not only the bare necessities of life such as adequate nutrition, clothing, and shelter but also facilities for reading, writing, expressing oneself, and interacting with other human beings without fear (Mullin v And'r, Union Territory of Delhi, 1981).

When algorithms model and predict human behavior, there are important implications for the dignity of the individuals targeted. Modelling of human behavior includes use cases where the intent is either to predict or understand the activities, motivations, or proclivities of human beings. This is true even for cases where the intent is not to model human behavior but the clear implication is on decisions taken regarding human beings, due to systemic factors involved in data collection and labelling, use of algorithms, and impact of inferences, etc. As an individual's data is manipulated and formatted to extract a pattern about that individual's world, the individual or their data no longer exists for itself (Cheney-Lippold, 2017), but are massaged into various categories. Amoores terms this a "data-derivative", which is an abstract conglomeration of data that continuously shapes our futures (Amoores,

2011). Cheney-Lippold argues that algorithmic agents create identities for us on their own terms, rarely with input from the subjects of the algorithm itself (Cheney-Lippold, 2017) and terms this construction a measurable (a data equivalent of Weber's ideal type) construct of conceptual purity that does not occur in reality (Cheney-Lippold, 2017). Moreover, Rouvroy argues that the operation of the algorithm in terms of mathematical precision ignores the embodied individual and replaces him with a datafied substrate that can in no way capture the complexities of his character (Rouvroy, 2013). This leads to mathematical conclusions on the features of a certain group that might not reflect reality. Yet, the datafied substrate, replete with assumptions compounded by hidden layers, is used for making targeted decisions.

These ramifications are amplified in the case of minorities and other vulnerable communities. Algorithmic discrimination has been a concern among both legal experts and technologists for some time. Hao explains three phases at which some form of algorithmic bias might play out (Hao, 2019). The first stage comes with the framing of the problem. As soon as developers create a deep-learning model, they decide what output they want the model to provide and the rules needed to achieve this output. However, as discussed earlier, notions of "credit-worthiness", "recruitability", "suspicious", or "at risk" are often subject to cognitive bias. This makes it difficult to devise screening algorithms, which fairly portray society and the conglomeration of identities, and power asymmetries that define it (Basu, 2019).

The second stage is the data collection phase. As we saw with the predictive policing setup in Delhi, often data does not adequately represent reality. As crime rates are determined based on the number of calls that come into the Delhi Police call center, the quality of the dataset is highly dependent on how seriously the receiver takes each call (Marda & Narayan, 2020a). Calls from women from lower socio-economic groups

alleging sexual violence are often not taken seriously (Marda & Narayan, 2020a). A related problem is that datasets that are well curated and readily available are often very limited. For example, the data used for Natural Language Processing Systems for Parts of Speech (POS) tagging in the US come from popular newspapers such as The Wall Street Journal. However, accuracy of these datasets would decrease if the speech used by Wall Street Journal writers were applied to individuals or ethnic minorities who speak with a very different style (Blackwell, 2015).

The final stage is that of data preparation, where the developer selects the parameters which they want the algorithm to consider. For example, when determining credit-worthiness, the candidate's type of employment might be a parameter. It could be argued that someone working in the informal economy may be less likely to financially sustain themselves and thus would be deemed less credit-worthy. However, many individuals working in the informal economy in India are from lower caste communities (Kar, 2018a). Thus, working in the informal economy is an ostensibly neutral proxy for discriminating against a specific caste, thereby violating the right to equality when the data is being sorted during the machine learning process (Prince & Schwarcz, 2020).

The right to equality has been enshrined in several international human rights instruments and into the Equality Code of the Indian Constitution. The dominant approach to interpreting this right appears to focus on the grounds of discrimination in Article 15(1), thereby eschewing unintentional discrimination and disparate impact on certain communities. However, as Bhatia highlights (Bhatia, 2016), a few cases have considered indirect discrimination to some extent – an approach that is critical in the case of data-driven decision-making. Hence, we articulate the specific question on evaluating potential impact on minority groups, so that developers think of the potentially negative consequences of supposedly well-intentioned decisions.

Guiding Questions

The following questions help guide regulations on agency, dignity, and non-discrimination:

- Is the AI solution modelling or predicting human behavior?
- Is the AI solution likely to impact individuals or communities, in particular the minority, protected, or at-risk groups?

Recommended Regulatory Strategy

If AI is modelling or predicting human behavior, the state must be compelled to justify why this is necessary and proportionate to the objective. This justification must mandatorily be provided by any entity choosing to apply AI for this purpose, and must be enforced through either legislation or executive order. If a private sector actor such as Staqu is involved in partnership with the government, it must go through a process of accreditation, which should be determined by a co-regulatory body. All projects must also go through a mandatory impact assessment that considers the possibility of disparate impact or proxy discrimination. This must be mandated through co-regulatory guidelines framed by the government in consultation with private sector actors. We believe that a co-regulatory framework with regular consultations works best if a private sector actor is involved with the technology, as the government alone might not fully understand the implications of this technology. We also recommend that the private sector actor not be involved with the final decision. For instance, with credit rating, a number of private sector firms are involved in crunching data from the traditionally financially underserved and predicting their behavior. However, the final decision to sanction or reject a loan must be taken by a loan officer from a bank.

Safety, Security, and Human Impact

The fundamental principle that guides regulatory decisions in this case is that of safety, security, and human impact. Where the use of AI has the potential for direct, adverse, or large-scale human impact, greater regulatory intervention is required. In the Berkman-Klein study, safety and security of AI systems are present in 81% of documents espousing ethical AI (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020). Therefore, the following broad questions need to be asked:

- Is there either a high likelihood or high severity of potential adverse human impact of the AI solution?
- Can the likelihood or severity of adverse impact be reasonably ascertained with existing scientific knowledge?

While we acknowledge that both likelihood and severity of impact, and the risks posed therein, are contextual, we believe that certain trends are worth noting. When AI systems model human behavior, it is much more likely to lead to an impact on the human beings in question, or those who may be seen as belonging to the same group or category by the algorithm. An AI solution that could cause greater harm if applied erroneously, such as one deployed for predictive policing, should be subject to more stringent standards, audits, and oversight than an AI solution designed to create a learning path for a student in the education sector. There could be cases where the behavior being modelled is not human, yet it could lead to significant human impact. For instance, an AI system that makes predictions about weather or environmental factors does not model human behavior but could be used to make assessments that directly impact human beings.

When considering the impact, it is imperative to look at both the severity and likelihood of the adverse impact. A high “likelihood” of harm indicates a high probability of the human rights, quality of life, and core value clusters being negatively impacted due to multiple pre-deployment factors, such as corrupted data sets or lack of awareness among users. Scale of harm indicates the extent of impact, which is determined by factors such as number of individuals impacted, while severity of harm can be determined by aspects such as clamping down on civil liberties or causing socio-economic distress.

In some cases, the likelihood of the adverse impact on human beings may be low, yet in the remote eventuality that it does lead to an adverse impact, its severity could be very high. For instance, the use of autopilot systems in aircraft navigation or in controlled trials where the number of people impacted are limited. The attention to both aspects of risk is essential, as often justifications for risky systems are based on low likelihood. However, even in cases where there is low likelihood of human harm, if the severity is high enough, it may still augur for greater regulatory scrutiny.

In situations where the likelihood or severity of harm cannot be reasonably ascertained, we recommend adopting the precautionary principle from environmental law and suggest that the solution not be implemented until scientific knowledge reaches a stage where it can reasonably be ascertained (Kriebel, et al., 2001).



Regulatory Strategy

The following table contains a list of possible impact scenarios and regulatory strategies

Outcome	Explanation Of Outcome	Recommended Regulatory Strategy
A) High Likelihood, High Severity	Scenarios where the state is involved in predicting human behavior (predictive policing/credit rating/predicting school dropouts) but training data is incomplete and a thorough impact assessment has not been conducted.	Ban or proscribe until underlying issues are solved to reduce likelihood of harm. If likelihood or severity cannot be gauged, then the solution must not be deployed.
B) Low Likelihood, High Severity	Scenarios where training data is robust but individuals relying on use case (flood prediction, crop price forecasting) may face dire economic consequences if solution works incorrectly.	State run human rights impact assessment that externally verifies compliance.
C) High Likelihood, Low Severity	Possible in pilot cases where data, methodology, and funding are not yet clear and safeguards have not been appropriately devised, or where AI is not directly impacting civil liberties or socio-economic rights (traffic management).	Strong redressal mechanisms that enable even one impacted individual to receive compensation, particularly if the initial estimation of severity is too low.
D) Low Likelihood, Low Severity	Where data is robust, methodology, troubleshooting, and outreach have been clearly devised, and use case is not directly impacting civil liberties or socio-economic rights.	Possible regulatory forbearance with strong industry-driven codes for standardization, evaluation, and redressal if private sector is involved.

Table 2: Impact thresholds



Accountability, Oversight, and Redress

Background and Explanation

This principle attempts to grapple with two challenges to fostering accountability. The first challenge lies in the delegation of human decision making at some level to an algorithm, which creates an algorithmic “black box” through which inputs are processed and outputs are generated (Pasquale F., 2015). A certain level of transparency is key to fostering accountability frameworks for algorithmic decision-making. Any algorithmic decision-making framework in the public sector should reasonably be able to explain its decision to anyone impacted by its working. However, there may be a trade-off between the capacity or complexity of a model and the extent to which it can render a reasonably understandable explanation (Oswald, 2018).

Retrospective adequation is a legal standard we propose to promote algorithmic accountability (Sinha & Mathews, 2020). Essentially, this means that whenever inferences from machine learning algorithms influence decision making in public functions, they can do so only if a human agent is able to look at the existing data and discursively arrive at the same conclusion. Unlike the right to explanation under the General Data Protection Regulation (GDPR), which only includes “meaningful information about the logic involved, as well as the significance of the envisaged consequences of processing”.¹⁰ As opposed to the case of retrospective adequation, it does not tell us how an inference has been reached. This approach essentially draws from standards of due process and accountability evolved in administrative law, where decisions taken by public bodies must be supported by recorded justifications. Since the Maneka Gandhi vs Union of India judgment in

1978, the Supreme Court of India has clearly espoused the idea of both procedural and substantive procedural fairness. A further extension of this principle is the need for administrative authorities to record reasons to exclude or minimize arbitrariness (A Vedachalal Mudaliar v State of Madras, 1952). In some jurisdictions such as the UK and US, there are statutory obligations that require administrative authorities to give reasoned orders.¹¹ While there is no such corresponding statutory provision in India, the case law is fairly instructive in imposing similar obligations of quasi-judicial authorities (Travancore Rayons v Union of India, 1971; Siemen Engineering and Manufacturing Co. of India v Union of India, 1976). As Pasquale argues, explainability is important because reason-giving is intrinsic to the judicial process and cannot be jettisoned on account of algorithmic processing (Pasquale, F.A., 2017). The same principles equally apply to all administrative bodies, as it is a well-settled principle of administrative law that all decisions must be arrived at after a thorough application of mind. Much like a court of law, these decisions must be accompanied by reasons to qualify as a “speaking order”. Where the administrative decisions are informed by an algorithmic process opaque enough to prevent this, the next logical question is whether a system can be built in such a way that it flags relevant information for independent human assessment to verify the machine’s inferences. Only then will the requirements of what we call a speaking order be in any position to be satisfied.

Our assessment of opportunity for human supervision is based on the idea that where inferences are inherently opaque, they must provide sufficient information about the model and data analyzed, such that a human supervisor must be in a position to apply analogue

10. Art.15 GDPR

11. Section 12 of the (UK) Tribunals and Enquiries Act, 1958; Section of the (US) Federal Administrative Procedural Act, 1946

modes of analysis to the information available in order to conduct an independent assessment. For instance, where AI systems are used to detect hate speech for takedown from online platforms, it is possible to make available the inferences to a human supervisor who can apply her mind independently to the speech in question based on legal rules and standards on hate speech and relevant contextual information.

The increased role of the private sector in designing and deploying AI systems poses a challenge. As established earlier, there remains no clear threshold for demarcating public functions with private ones. With an increase in for-profit private actors playing a role in the discharge of functions that may be public, a liability mechanism that enables redress for adversely impacted individuals needs to be thought through. A potential thorny issue may be the proprietary nature of the source code, which the private sector developer may not want to share. This makes it imperative to think around unique regulatory interventions to constrain the private sector actor within the framework of the rule of law. This is particularly significant for start-ups, such as those involved in credit rating, who want to do “social good” but do not have the financial resources or bandwidth to create their own voluntary compliance strategy. Therefore, regulatory certainty that clearly demarcates scope of activity, liability, and evaluation metrics for private sector actors is vital.

The following questions help determine accountability, oversight, and redress:

- To what extent is the AI solution built with human-in-the-loop supervision prospects?
- Are there reliable means for retrospective adequation?
- Is the private sector partner involved with either the design of the AI solution, its deployment, or both?

Smart Regulation Strategy

Since an empirical mapping of the potential loopholes in AI implementation across India’s socio-economic demographics does not exist, all AI solutions must be built with human-in-the-loop supervision. Essentially, this means that while AI can aggregate and analyze data on a certain issue, the final decision will need to be taken by a human being. As our case studies showed, human bias in decision-making was prevalent well before machine learning came into the picture. However, human beings can be questioned, engaged with, and held accountable through legal proceedings – something that cannot be done with an AI system. In addition, human beings also retain the flexibility to make broader policy interventions. For example, if it is observed that crime rates are higher among a certain community, instead of merely trying to stamp out crime, a human being might try to identify the root cause of the crime, which might lie in higher rates of unemployment or poverty in the area. Therefore, they may look to intervene by devising social welfare programs instead of merely conducting enhanced surveillance. As such, human-in-the-loop must be made mandatory through top-down legislation.

Retrospective adequation is necessary for imposing accountability on AI systems discharging public functions and impacting citizens’ rights. We recommend the evolution of technical standards from the private sector actors operating in India, which are then discussed and affirmed by a co-regulatory body such as the Bureau of Indian Standards.

If a private sector actor is involved with the design or deployment of the AI solution, then it must be first considered whether the activity in question falls within a reasonable and contextual understanding of a “public function”. It is clear that private sector actors should

not deploy solutions when it comes to three core governmental functions: foreign relations, any form of violence or provision of security, and legislation. This essentially means that once the final decision is taken, any follow-up action must be decided and acted upon by a government entity.

Actors such as Staqu are involved in the design and development of the AI solution, even though the police implement the recommended outcome. Moreover, cases of public service delivery that have clear implications for the realization of the right to life could be considered public functions. Either the state or private actor must be held liable if rights are violated in the process. To encourage private actors to participate, the state may choose to soak up some of the liability for damages. However, clear mechanisms for assignment of liability must exist – something that was not done for Microsoft’s partnership with the government of Karnataka. In such cases, consistent obligations must be imposed on the private sector. To this end, we recommend:

- Clearly drafted contracts with private sector developers that specify modes of liability, nature, and frequency of audits and impact assessments, as well as clarification that their source code and training data may need to be made public if the algorithmic decision-making is challenged in a court of law.
- Internal decision-making processes within the organization must be scrutinized for conformity with constitutional standards and human rights.
- The organization must ensure that they will not interfere with core government decision-making processes, such as deciding when to use violence in the interest of public order.

- In cases where private actors are involved with any function that violates civil and political or socio-economic rights, and an aggrieved individual(s) challenges the violation in a court of law, the court must treat this as a “public function” and hold the private sector actor to the same level of scrutiny as the government. If the government wants to shield the private actor from this liability, then it must be explicitly stated in the contract. These contracts must also be made public.
- That the private sector actor provides the needed capacity building to public sector actors to ensure they can understand the functioning and outputs of the system.

Privacy and Data Protection

Explanation

It is often argued that for emerging economies, the right to privacy should take a backseat to development. However, as we have highlighted in this paper, the poor and vulnerable are the most likely to have their civil liberties infringed by data-driven decision-making. When affirming the right to privacy as a fundamental right, the Indian Supreme Court strongly rebutted this, arguing that civil and political rights are important for every individual regardless of income (K. Puttaswamy v Union of India, 2017). They also affirmed that placing socio-economic rights over civil and political rights has been done away with by constitutional courts. Since this judgement in 2017, India has sought to formulate a data protection law – tabling a bill in Parliament in December 2019 (Basu & Sherman, 2020). While the obligations on private data processors in the bill are similar, it does some disservice to individual rights by granting the government a wide range of exceptions.



Regulatory interventions for predictive policing

Value	Questions	Predictive Policing	Regulatory Intervention
<p>Agency</p>	<p>Is adoption of the solution mandatory?</p>	<p>Mandatory for all police officers depending on the decision made by police chief functionaries and mandatory for individuals that the police decide to use the solution on.</p>	<p>Regular consultation and feedback from all levels within the police hierarchy, in particular officers who directly engage with victims on the ground and the public.</p> <p>Notice to individuals when a decision about them has been taken using an AI system.</p> <p>Human rights impact assessment.</p>
	<p>Does the solution allow for end-user control?</p>	<p>Yes, as the police officer using it is the end user.</p>	<p>N/A</p>
	<p>Is there a vast disparity between the primary user and the impacted party?</p>	<p>Yes, between police officers and suspected criminals.</p>	<p>Mandatory certification for all police officers working both with the algorithm and implementing it on the ground (through notification).</p> <p>Statistical standards for accuracy.</p> <p>Evidentiary weight of decisions informed by an AI system.</p>

Table 3a: Regulatory interventions for predictive policing

Equality, Dignity, and Non-Discrimination	Is the AI solution modelling or predicting human behavior?	Modelling criminality.	Needs assessment from the decision maker on why modelling human behavior is proportionate to the objective of reducing crime and also demonstrating why no other reasonable alternatives exist.
	Is the AI solution likely to impact minority, protected, or at-risk groups?	Possible disparate impact.	Awareness, sensitization, and creation of grievance redressal mechanisms and anti-discrimination regulations protecting vulnerable groups.
Safety, Security, and Human Impact	Is there a high likelihood or high severity of potential adverse human impact as a result of the AI solution?	Possible high likelihood and high severity, unless data collection practices are improved.	Proscription of solution until data curation and analysis is improved and standardized. The use of the system should be guided by the principles of necessity, proportionality, and least intrusive means. Compliance with international security standards.
	Can the likelihood or severity of adverse impact be reasonably ascertained with existing scientific knowledge?	Yes, through empirical research.	Government and the private sector should undertake regular empirical assessments of potential impact.

(Cont.) Table 3a: Regulatory interventions for predictive policing

Accountability, Oversight, and Redress	To what extent is the AI solution built with "human-in-the-loop" supervision prospects?	Human-in-the-loop exists.	
	Are there reliable means for retrospective adequation?	No publicly available information.	The private actor involved should mandatorily demonstrate possibility of retrospective adequation.
	Is the private sector partner involved with either the design of the AI solution, its deployment, or both?	Yes.	Contract as described above. Final implementation of the decision should continue to be done by the police.
Privacy and Data Protection	Does the AI solution use personal data, even in anonymized form?	Yes.	Any data collection must comply with a national data protection law that clearly separates personal and non-personal data.

(Cont.) Table 3a: Regulatory interventions for predictive policing

Regulatory interventions for credit rating

Value	Questions	Predictive Policing	Regulatory Intervention
<p>Agency</p>	<p>Is adoption of the solution mandatory?</p>	<p>Optional for loan-providers from banks. They can potentially switch to a credit rating company that does not use AI.</p>	<p>Banks should have an internal regulatory strategy on the adoption of AI.</p> <p>Human rights impact assessment.</p>
	<p>Does the solution allow for end-user control?</p>	<p>Yes, as the company/ bank engaging in credit rating is the end-user.</p>	<p>N/A</p>
	<p>Is there a vast disparity between the primary user and the impacted party?</p>	<p>Yes, there is a disparity between those generating the scores and those they are scoring.</p>	<p>Self-regulation: Loan officers and credit rating companies should communicate clearly to potential candidates the decision-making process, how AI is being used, and possible implications.</p>
<p>Equality, Dignity, and Non-Discrimination</p>	<p>Is the AI solution modelling or predicting human behavior?</p>	<p>It is determining “credit-worthiness”.</p>	<p>Mandatory needs assessment from bank clarifying why algorithmic decision-making is more accurate than traditional credit scoring methods, as well as full transparency on data being used and curation methods.</p>
	<p>Is the AI solution likely to impact minority, protected, or at-risk groups?</p>	<p>Possible disparate impact.</p>	<p>Awareness, sensitization, training and creation of grievance redressal mechanisms targeting vulnerable groups.</p>

Table 3b: Regulatory interventions for credit rating

Safety, Security, and Human Impact	Is there a high likelihood or high severity of potential adverse human impact as a result of the AI solution?	Possible high likelihood and high severity.	Mandatory pilot projects and standardization of data curation practices certified by a co-regulatory committee.
	Can the likelihood or severity of adverse impact be reasonably ascertained with existing scientific knowledge?	Yes.	
Accountability, Oversight, and Redress	To what extent is the AI solution built with "human-in-the-loop" supervision prospects?	Human-in-the-loop exists.	
	Are there reliable means for retrospective adequation?	No publicly available information.	Retrospective adequation should comply with Indian credit regulations.
	Is the private sector partner involved with either the design of the AI solution, its deployment, or both?	Both.	Contract as described above. If the private sector partner is a start-up, the state may choose to cushion some of the liability. Final decision must be independently taken by the bank sanctioning the loan.
Privacy and Data Protection	Does the AI solution use personal data, even in anonymized form?	Yes.	Any data collection must comply with a national data protection law that clearly separates personal and non-personal data.

(Cont.) Table 3b: Regulatory interventions for credit rating

Regulatory interventions for AI in agriculture

Value	Questions	Agriculture	Regulatory Intervention
Agency	Is adoption of the solution mandatory?	No, farmers may opt out.	Pros and cons of adopting the solution should be clearly communicated in an understandable format to the farmer (self-regulation).
	Does the solution allow for end-user control?	Yes, the farmer using the solution is the end-user.	N/A
	Is there a vast disparity between the primary user and the impacted party?	No, the farmer is the end-user and feels the impact of the solution.	A co-regulatory consultative body should be set up to organize regular consultations between the users and the developers of the project.
Equality, Dignity, and Non-Discrimination	Is the AI solution modelling or predicting human behavior?	It is modelling crop patterns and weather data.	
	Is the AI solution likely to impact minority, protected, or at-risk groups?	No, while there may be a negative impact, it is unlikely to specifically impact minorities.	All farmers may not equally benefit from the app. Government and private sector partners must mandatorily provide training, set up a pre-requisite infrastructure to the extent possible, and also study trends on why certain farmers may not be benefitting.

Table 3c: Regulatory interventions for AI in agriculture

Safety, Security, and Human Impact	Is there a high likelihood or high severity of potential adverse human impact as a result of the AI solution?	Depending on the quality of the data curated, there is possible low likelihood and low severity.	Mandatory pilot projects and standardization of data curation practices as certified by a co-regulatory committee.
	Can the likelihood or severity of adverse impact be reasonably ascertained with existing scientific knowledge?	Yes.	The private sector partner could publish research on preliminary scientific studies (voluntarism).
Accountability, Oversight, and Redress	To what extent is the AI solution built with "human-in-the-loop" supervision prospects?	Unclear.	More public information about the working of the app should be disclosed to the public and to the farmers concerned.
	Are there reliable means for retrospective adequation?	No publicly available information.	The private sector partner should be able to provide retrospective adequation for all decisions.
	Is the private sector partner involved with either the design of the AI solution, its deployment, or both?	Both.	There needs to be a contract clearly imposing liability on the private sector partner in case of negligence. If the private sector partner is a start-up, the state may choose to cushion some of the liability.
Privacy and Data Protection	Does the AI solution use personal data, even in anonymized form?	Yes.	Any data collection must comply with a national data protection law that clearly separates personal and non-personal data.

(Cont.) Table 3c: Regulatory interventions for AI in agriculture

Conclusion

The application of regulatory interventions to use cases brought up a number of similarities. While predictive policing is a core government function that could involve violence further down the line, the *modus operandi*, and therefore the potential threats to core constitutional values are similar to those in credit rating. The fundamental difference between these two use cases and the agricultural case study is that these involved two sets of human beings – one group being in a position of power that is attempting to predict how less powerful human beings will act. Thus, the regulatory interventions needed to optimally govern AI stem from those necessary to remedy structural injustices in society. The danger, however, in both India and other parts of the world, stems from technological solutionism, which assumes that existing societal fissures can be occluded through data-driven decision making. The reality is quite different, with data-driven decision making needing to adapt the same values that were required to fairly govern society in a pre-AI world. This is compounded by a lack of effective public oversight and consultation of both policymaking and technological implementation. There are no publicly scrutable external impact assessments post-deployment or publicly available empirical socio-economic assessments prior to deploying the solution.

Our paper establishes a framework for adapting these values through a series of questions that identify critical junctures at which core constitutional values and human rights may be at threat due to algorithmic decision-making. Our framework is by no means exhaustive and is meant to be read as a set of guidelines for decision makers and technologists looking to devise their own set of frameworks. The set of regulatory tools mapped out by Freiburg (2010) may remain relevant and need to be applied across contexts – often in response to knowledge that may be gained as the AI solution is implemented, evaluated, and adapted.

The five sets of values that we felt merited protection: (1) agency; (2) equality, dignity, and non-discrimination; (3) safety, security, and human impact; (4) accountability,

oversight, and redress; and (5) privacy and data protection, were selected not only from a study of India's constitutional fiber but also through an assessment of AI policy instruments released by a variety of stakeholders around the world. As such, we feel that our framework – although researched and developed in an Indian context – applies across emerging economies who desire to improve the government's role in public service delivery while still mitigating negative impacts.

A core challenge continues to be the complex question of the involvement of the private sector in functions that have traditionally been the government's prerogative, and often those that have implications for fundamental rights. One of the most important recommendations of our paper centers around the need to hold the private sector accountable in these instances through uniformly worded contracts that adequately impose liability along with the delegation of any responsibility. However, given the lack of government capacity to entirely identify, design, and deploy an AI-driven solution, some regulatory room must be given to these actors to innovate.

Appropriate regulation therefore does not fit neatly into the division of the modes of hierarchical regulation, co-regulation, and self-regulation. A smart regulatory strategy would require a combination of all three.

Going forward, we feel the need for more empirical assessment of use cases in emerging economies, as much of the literature, both on the technology and regulatory frameworks, are devised in a Western context and therefore not entirely applicable to emerging economies. That said, our paper shows that algorithmic decision-making is becoming more commonplace in emerging economies. Through a close analysis of the information gained from these empirical assessments and a strong commitment to the values described, we believe that adequate ex ante regulation can mitigate harms while also enabling the realization of prospects for social good.

Acknowledgements

This paper was shaped by several helpful conversations with practitioners and scholars who were incredibly generous with their time. We would like to thank Malavika Raghavan, Srikara Prasad, Vidushi Marda, Sushant Kumar, and Anita Srinivasan. The paper also benefited from feedback received after presentations at the Tamil Nadu e-governance agency and Microsoft Research in Bengaluru. We were honored to be a part of the excellent cohort and benefited greatly from the support offered by colleagues involved with this Association of Pacific Rim Universities (APRU) project.

This paper was greatly improved by edits and feedback provided by Vipul Kharbanda, Nikhil Dave, and Divij Joshi. We would also like to thank Nikhil Dave for some excellent research assistance on this paper. All errors remain our own.

References

A Vedachalal Mudaliar v State of Madras, AIR Mad. 276 (1952)

Academic Center of Law and Business v Minister of Finance, Isr. (2006, Aug 20)

Agarwal, S. (2018, February 20). Internet users in India expected to reach 500 million by June: IAMAI. Retrieved from The Economic Times: <https://economictimes.indiatimes.com/tech/internet/internet-users-in-india-expected-to-reach-500-million-by-june-iamai/articleshow/63000198.cms>

Agrawal, A. (2020, May 1). Lockdown Extension: Aarogya Setu Mandatory for All Employees and in Containment Zones. Retrieved from MEDIANAMA: <https://www.medianama.com/2020/05/223-coronavirus-lockdown-extended-by-2-weeks-country-divided-into-red-orange-and-green-zones/>

Alabama v White, 496 U.S. 325 (1990)

Amoore, L. (2011). Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times. SAGE journal, 24, 27. Retrieved from <https://journals.sagepub.com/doi/10.1177/0263276411417430>

Arun, C. (2019). AI and the Global South: Designing for Other Worlds. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), the Oxford Handbook of Ethics of AI. Oxford University Press. Retrieved from <https://ssrn.com/abstract=3403010>

Ayres, I., & Braithwaite, J. (1992). Responsive Regulation. Oxford University Press.

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. 104 California Law Review 671.

Barrows v Jackson, 252, U.S. (1953)

Basu, A. (2019, October 12). We Need a Better AI Vision. Fountainink. Retrieved from Fountain Ink: <https://fountainink.in/essay/we-need-a-better-ai-vision->

Basu, A., & Hickok, E. (2018). Artificial Intelligence in the Governance Sector in India. India: The Centre for Internet and Society. Retrieved from <https://cis-india.org/internet-governance/ai-and-governance-case-study-pdf>

Basu, A., & Pranav, M. (2019, July 21). What is the problem with 'Ethical AI'? An Indian Perspective . Retrieved from The Centre for Internet and Society: <https://cis-india.org/internet-governance/blog/what-is-the-problem-with-2018ethical-ai2019-an-indian-per>

Basu, A., & Sherman, J. (2020, January 23). Key Takeaways from India's Revised Personal Data Protection Bill. Lawfare.

Berg, N. (2014, June 25). Predicting Crime, LAPD style. Retrieved from The Guardian: <https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report>

Bhatia, G. (2016). Retrieved from Indian Constitutional Law and Philosophy: <https://indconlawphil.wordpress.com/tag/indirect-discrimination/>

Bhatia, G. (2017). Equal moral membership: Naz Foundation and the refashioning of equality under a transformative constitution. *Indian Law Review*, 115-144.

Bhattacharya, P. (2016, December 5). 88% of households in India have a mobile phone. Retrieved from Livemint: <https://www.livemint.com/Politics/kZ7j1NQf5614UvO6WURXfO/88-of-households-in-India-have-a-mobile-phone.html>

Black, J. (2001). Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World 54 *Current Legal Problems* (Vol. 54). *Current Legal Problems*.

Blackwell, A. F. (2015). Interacting with an Inferred World: The Challenge of Machine Learning for Humane Computer Interaction". *Proceedings of the Fifth Decennial Aarhus Conference on Critical Alternatives*, 179.

Braithwaite, J. (2000). The New Regulatory State and the Transformation of Criminology. *British Journal of Criminology*, 40, 222-38. <http://doi:10.1093/bjc/40.2.222>

Capital Float (2020). Retrieved from Capital Float: <https://capitalfloat.com/>

Capon, N. (1982). Credit Scoring Systems: A Critical Analysis. *Journal of Marketing*, 46(2), 82-91. Retrieved from <https://www.jstor.org/stable/3203343>

Charyulu, D. K., Shyam, D. M., Wani, S. P., & Raju, K. (2017). Rythu Kosam: Andhra Pradesh Primary Sector Mission. Coastal Andhra Region Baseline Summary Report.

ICRISAT Development Center. Retrieved from <http://111.93.2.168/idc/wp-content/uploads/2018/01/IDC-Report-No-13-Rythu-Kosam.pdf>

Chatterjee, S., & Kapur, D. (2016). Understanding Price Variation in Agricultural Commodities in India: MSP, Government Procurement, and Agriculture Markets. India Policy Forum. Retrieved from <http://www.ncaer.org/events/ipf-2016/IPF-2016-Paper-Chatterjee-Kapur.pdf>

Cheney-Lippold, J. (2017). We Are Data: Algorithms and the Making of Our Digital Selves. NYU Press.

Christiansen, E. C. (2011, January 1). Transformative Constitutionalism in South Africa: Creative Uses of Constitutional Court Authority to Advance Substantive Justice. SSRN. Retrieved from <https://ssrn.com/abstract=1890885>

Chugh, B., & Raghavan, M. (2019, June 18). The RBI's proposed Public Credit Registry and its implications for the credit reporting system India. Retrieved from Dvara Research: <https://www.dvara.com/blog/2019/06/18/the-rbis-proposed-public-credit-registry-and-its-implications-for-the-credit-reporting-system-in-india/>

Citron, D. K., & Pasquale, F. A. (2014). The Scored Society: Due Process for Automated Predictions. Washington Law Review, 14, 89.

Commission, E. (n.d.). Ethics Guidelines for trustworthy AI. Retrieved from European Commission: <https://ec.europa.eu/futurium/en/ai-alliance-consultation>

Common Cause. (2018). Status of Policing in India Report 2018: A Study of Performance and Perceptions. Common Cause & Lokniti - Centre for the Study Developing Societies (CSDS). Retrieved from <https://www.commoncause.in/pdf/SPIR-2018-c-v.pdf>

Corbett-Davies, S. (2017). Algorithmic Decision-making and the Cost of Fairness. Stanford University. Retrieved from http://www.antonioacasella.eu/nume/Corbett-Davies_2017.pdf

Das, S. (2017, March 21). How Predictive Analytics Helps Indian Police Fight Crime. Retrieved from <http://www.computerworld.in/feature/how-predictive-analytics-helps-indian-police-fight-crim>

Department for Promotion of Industry and Internal Trade. (2018). Report of Task Force on Artificial Intelligence. Government of India. Retrieved from <https://dipp.gov.in/whats-new/report-task-force-artificial-intelligence>

Desai, K. (2019, March 31). Now Police Use Apps to Catch a Criminal. Retrieved from Times of India: <https://timesofindia.indiatimes.com/home/sunday-times/now-police-use-apps-to-catch-a-criminal/articleshow/68649118.cms>

Deshpande, R. S. (2002, June 29). Suicide by Farmers in Karnataka Agrarian Distress and Possible Alleviatory Steps. *Economic and Political Weekly*, pp. 2601-2604. Retrieved from http://shreeindia.info/rsdeshpande.com/wp-content/uploads/2014/03/Suicide_by_Farmers_in_Karnataka.pdf

Doekler, A. (2010). Self-regulation and Co-regulation: Prospects and Boundaries in an Online Environment. Master of Law thesis, University of British Columbia. Retrieved from <https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0071207>

Express Web Desk. (2017, October 27). Karnataka govt inks MoU with Microsoft to use Artificial Intelligence for digital agriculture. Retrieved from *The Indian Express*: <https://indianexpress.com/article/india/karnataka-govt-inks-mou-with-microsoft-to-use-artificial-intelligence-for-digital-agriculture-4909470/>

Federal Trade Commission Staff. (2009). Report on Self-regulatory Principles for Online Behavioral Advertising. Retrieved from <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-beh>

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020, January 15). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research. Retrieved from <https://ssrn.com/abstract=3518482>

Francis Coralie Mullin v UT of Delhi, *AIR 746 (1981)*.

Freeman, J. (2000). The Private Role in Public Governance. *NYULR*, 75, 543, 547, 651–53.

Freiberg, A. (2010). Restocking the Regulatory Tool-kit. Dublin. Retrieved from <http://www.regulation.upf.edu/dublin-10-papers/111.pdf>

Ganguly, S. (2020, March 31). Gurugram-based Start-up Staqu Has Notified AI-powered JARVIS to Battle Coronavirus. Retrieved from *Your Story*: <https://yourstory.com/2020/03/gurugram-ai-startup-staqu-jarvis-coronavirus>

Gateway, F. (2019, January 29). India: Reserve Bank of India Is Working on Public Credit Registry to Improve Access to Micro Credit. Retrieved from *FinDev Gateway*: <https://www.findevgateway.org/news/india-reserve-bank-india-working-public-credit-registry-improve-access-micro-credit>

Goudarzi, S., Hickok, E., & Sinha, A. (2018). AI in Banking. India: The Centre for Internet and Society. Retrieved from <https://cis-india.org/internet-governance/files/ai-in-banking-and-finance>

Government of India. (2006). Notification. India. Retrieved from <https://rbidocs.rbi.org.in/rdocs/Content/PDFs/69700.pdf>

Government of India. (2019). Data "Of the People, By the People, For the People."

Government of India. (2019). Draft National E-Commerce Policy. Retrieved from https://dipp.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf

Guihot, M., Matthew, A. F., & Suzor, N. P. (2017). Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence. *VJETL*, 20(2), 385, 429. Retrieved from http://www.jetlaw.org/wp-content/uploads/2017/12/2_Guihot-Article_Final-Review-Complete_Approved.pdf

Gunningham, N., & Sinclair, D. (2017). Smart Regulation. In P. Drahos (Ed.), *Regulatory Theory: Foundations and Applications* (p. 115). ANU Press.

Gurumurthy, A., & Bharthur, D. (2019). Taking Stock of AI in Indian Agriculture. *IT for Change*. Retrieved from <https://itforchange.net/sites/default/files/1664/Taking-Stock-of-AI-in-Indian-Agriculture.pdf>

Hailbronner, M. (2017, November 22). Transformative Constitutionalism: Not Only in the Global South. *American Journal of Comparative Law*, 65(3), 527-556. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2777695

Haines, F. (2017). Regulation and Risk. In P. Drahos (Ed.), *Regulatory Theory, Foundations and Applications* (p.181). ANU Press.

Hao, K. (2019, February 4). This is how AI bias really happens—and why it's so hard to fix. Retrieved from MIT Technology Review: <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/>

Heerdt, J. (n.d.). Transform your data into intelligent action with Cortana Analytics Suite. Retrieved from Sogeti: <https://www.sogeti.nl/sites/default/files/Transform%20your%20data%20into%20intelligent%20action%20with%20Microsoft%20Cortana%20Analytics%20Platform.pdf>

Hood, C. C., & Margetts, H. Z. (2008). *The Tools of Government in the Digital Age*. Palgrave Macmillan.

IANS. (2016, June 9). Microsoft develop sowing app for Andhra Pradesh farmers. Retrieved from Financial Express: <https://www.financialexpress.com/industry/technology/microsoft-develop-sowing-app-for-andhra-pradesh-farmers/279171/>

IANS. (2017, December 19). #GoodNews: Indian Farmers Go the AI Way to Increase Crop Yields. Retrieved from the quint: <https://www.thequint.com/news/india/good-news-indian-farmers-use-ai-for-higher-crop-yields>

ICICI Bank. (2020, January 1). Artificial Intelligence in Loan Assessment: How does it Work? Retrieved from ICICI Bank: <https://www.icicibank.com/blogs/personal-loan/artificial-intelligence-in-loan-assessment-how-does-it-work.page?>

ICRISAT. (2017, January 9). Microsoft and ICRISAT's Intelligent Cloud Pilot for Agriculture in Andhra Pradesh Increase Crop Yield for Farmers. Retrieved from ICRISAT: <http://www.icrisat.org/microsoft-and-icrisats-intelligent-cloud-pilot-for-agriculture-in-andhra-pradesh-increase-crop-yield-for-farmers/>

ICRISAT. (2017, January 13). New Sowing Application Increases Yield by 30%. Retrieved from ICRISAT: <http://www.icrisat.org/new-sowing-application-increases-yield-by-30/>

ICRISAT. (n.d.). Microsoft CEO Speaks on Collaboration with ICRISAT. Retrieved from ICRISAT: <http://www.icrisat.org/microsoft-ceo-speaks-on-collaboration-with-icrisat/>

ICRISAT. (n.d.). Rythu Kosam. Retrieved from ICRISAT: <http://www.icrisat.org/tag/rythu-kosam>

Illinois v Gates, 462 U.S. 213 (1983)

Indian Express (2018, March 16). Why are India's Farmers Committing Suicide? Retrieved from Indian Express: <http://www.newindianexpress.com/nation/2018/mar/15/why-are-indias-farmers-committing-suicide-1787539.html>.

Jaggi, S. (2017). State Action Doctrine. Max Planck Encyclopedia of Comparative Constitutional Law. Retrieved from <https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e473>

Jessop, R. (2003). Governance and Metagovernance: On Reflexivity, Requisite Variety, and Requisite Irony. Sociology. Lancaster University. Retrieved from <https://www.lancaster.ac.uk/fass/resources/sociology-online-papers/papers/jessop-g>

Joshi, D. (2020, February 6). Welfare Automation in the Shadow of the Indian Constitution. Retrieved from Socio-Legal Review: <https://www.sociolegalreview.com/post/welfare-automation-in-the-shadow-of-the-indian-constitution>

K. Puttaswamy v Union of India (I) 10 SCC 1, 2017

Kahn, J. (2020, February 11). A.I. and tackling the risk of "digital redlining". Retrieved from Fortune: <https://fortune.com/2020/02/11/a-i-fairness-eye-on-a-i/>

Kaleidofin. (n.d.). About Us. Retrieved from Kaleidofin: <https://kaleidofin.com/about-us/>

Kannabiran, K. (2012). Tools of Justice: Non-Discrimination and the Indian Constitution. New York: Routledge.

Kar, S. (2018-a). Financializing Poverty: Labour and Risk in Indian Microfinance. Stanford University Press, 153.

Kar, S. (2018-b). Financializing Poverty: Labour and Risk in Indian Microfinance. Stanford University Press, 154.

- Khaitan, N. (2019, October 25). New Act UAPA: Absolute Power to State. Retrieved from Frontline: <https://frontline.thehindu.com/cover-story/article29618049.ece>
- Khaitan, T. (2009). Reading Swaraj into Article 15: A New Deal for the Minorities. NUJS Law Review.
- Khanikar, S. (2018). State Violence and Legitimacy in India, 321.
- Khilnani, S. (2004). The Idea of India. New Delhi: Penguin.
- Kleinstauber, H. J. (n.d.). Self-regulation, Co-regulation, State Regulation. Retrieved from <https://www.osce.org/fom/13844?download=true>
- Kriebel, D., Tickner, J., Epstein, P., Lemons, J., Levins, R., Loechler, E. L. . . . Stot, M. (2001). The Precautionary Principle in Environmental Science. Environmental Health Perspectives, 871-876.
- Kumar, A., Shukla, P., Sharan, A., & Mahindru, T. (2018). NationalStrategy-for-AI-Discussion-Paper. NITI Aaygo. Retrieved from https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf
- Langenbucher, K. (2020). Responsible A.I. Credit Scoring – A Legal Framework. 25 Euro. L. Rev. 1.
- Lending Kart (2020). Retrieved from Lending Kart: <https://www.lendingkart.com/>
- Lloyd Corp Ltd v Tanner*, 562, U.S. (1953)
- Loan Frame (2020). Retrieved from Loan Frame: <https://www.loanframe.com/>
- Maneka Gandhi v Union of India*, SCR (2) 621 (1978)
- Marda, V., & Narayan, S. (2020a). Data in New Delhi,s Predictive Policing System. Proceedings of ACM Conference on Fairness, Accountability, and Transparency. Barcelona, Spain, ACM, New York, NY, USA. USA. Retrieved from <https://doi.org/10.1145/3351095.3372865>
- Marda, V., & Narayan, S. (2020b). Data in New Delhi’s Predictive Policing System. Proceedings of ACM Conference on Fairness, Accountability, and Transparency, (p. 321). Barcelona, Spain. ACM, New York, NY, USA. USA. Retrieved from <https://doi.org/10.1145/3351095.3372865>
- Marda, V., & Narayan, S. (2020c). Data in New Delhi’s Predictive System. Proceedings of ACM Conference on Fairness, Accountability, and Transparency, (p. 322). Barcelona, Spain, ACM, New York, NY, USA. USA. Retrieved from <https://doi.org/10.1145/3351095.3372865>
- Mittelstadt, B. (2019, May 20). AI Ethics – Too Principled to Fail? Nature Machine Intelligence. Retrieved from Nature Machine Intelligence: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293

Mullin v And'r, Union Territory of Delhi, India, 2 S.C.R. 516, 518 (1981)

Nag, R. (2016, June 10). How Matrix Backed FinTech Startup Finomena is Disrupting the \$8 Bn Youth Loan Market. Retrieved from Inc 42: <https://inc42.com/startups/finomena/>

NASSCOM. (2018). Agritech In India – Maxing India Farm Output. Retrieved from NASSCOM: <https://www.nasscom.in/knowledge-center/publications/agritech-india-%E2%80%93-maxing-india-farm-output>

Nayak, N. D. (2015, May 3). Agricultural sector needs technological intervention to face challenges. Retrieved from The Hindu: <https://www.thehindu.com/news/national/karnataka/agricultural-sector-needs-technological-intervention-to-face-challenges/article7166263.ece>

NITI Aayog. (2018). National Strategy fir Artificial Intelligence. 33-34. Retrieved from http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

Oswald, M. (2018). Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. SSRN.

Palmer, S. (2008, July–October). Public Functions and Private Services: A Gap in Human Rights Protection. *International Journal of Constitutional Law*, 6(3-4), 585-60.

Partap Singh (Dr) v Director of Enforcement, Foreign Exchange Regulation Act, AIR SC 989 (1985)

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquale, F. A. (2017). Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. SSRN, 78. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3002546

Pearson, J. (2017). AI Could Resurrect a Racist Housing Policy. Retrieved from https://www.vice.com/en_us/article/4x44dp/ai-could-resurrect-a-racist-housing-policy

Pelaez, V. (2019). The Prison Industry in the United States: Big Business or a New Form of Slavery? *Global Research*. Retrieved from <https://www.globalresearch.ca/the-prison-industry-in-the-united-states-big-business-or-a-new-form-of-slavery/8289>

Pichai, S. (2018, June 7). AI at Google: Our Principles. Retrieved from Google: The Keyword: <https://www.blog.google/technology/ai/ai-principles/>

Pischke v Litscher, 178 F.3d 497, 500 (7th Cir. 1999)

Prince, A., & Schwarcz, D. (2020). Proxy Discrimination in the Age of Artificial Intelligence and Big Data. 105 Iowa Law Review 1257. Retrieved from <https://ssrn.com/abstract=3347959>

Randazzo, A. (2013). Can a Disruptive Fin-tech create a Mass Market for Savings and Investment in India? Retrieved from Kaleidofin: <https://kaleidofin.com/kaleidofin-can-a-disruptive-fin-tech-create-a-mass-market-for-savings-and-investment-in-india>

Ranger, C. (2018, November 13). Using machine learning to improve lending in the emerging markets. Retrieved from Harvard Business School - Technology and Operations Management: <https://digital.hbs.edu/platform-rctom/submission/using-machine-learning-to-improve-lending-in-the-emerging-markets/>

Rao, N. (2013). Three Concepts of Dignity in Constitutional Law. Notre Dame Law Review, 200.

Reddy, B. D. (2016, June 9). Microsoft, Icrisat develop new sowing app for farmers using AI and Azure cloud. Business Standard. Retrieved from https://www.business-standard.com/article/companies/microsoft-icrisat-develop-new-sowing-app-for-farmers-using-ai-and-azure-cloud-116060900752_1.html

Rouvroy, A. (2013). The End(s) of Critique: Data Behaviourism versus Due Process. In M. Hildebrandt, & K. De Vries, Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology.

Scherer, M. U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. Harvard Journal of Law & Technology, 29, 354, 357, 259.

Schulz, W. (2006). Final Report: Study on Co-regulation Measures in the Media Sector, Study for the European Commission. Directorate Information Society and Media . Retrieved from http://ec.europa.eu/avpolicy/docs/library/studies/coregul/final_rep_

Schulz, W., & Held, T. (2001). Regulated self-regulation as a form of modern government. Indiana University Press.

Scott, C. (2017a). The Regulatory State and Beyond. In Regulatory Theory, Foundations and Applications (p. 269). Australia: ANU Press.

Scott, C. (2017b). The Regulatory State and Beyond. In Regulatory Theory; Foundations and applications (pp. 269–270). ANU Press.

Sethia, A. (2015, March 21). "The BCCI Case on "Public Function" and Its Implications on Sports Governance. Retrieved from iconnectblog: <http://www.iconnectblog.com/2015/03/bcci-case-on-public-function/>

Sharma, S. (2018, July 9). How ISRO is helping in Uttar Pradesh Map and Predict Crime. Retrieved from Tech Circle: <https://www.techcircle.in/2018/07/09/how-isro-is-helping-uttar-pradesh-police-map-and-predict-crime/>

Sharma, V. (2017, September 23). Indian Police to be armed with Big Data Software to Predict Crime. Retrieved from The New Indian Express: <https://www.newindianexpress.com/nation/2017/sep/23/indian-police-to-be-armed-with-big-data-software-to-predict-crime-1661708.html>

Siemen Engineering and Manufacturing Co. of India v Union of India, AIR Sc 1785 (1976)

Singh, A., & Prasad, S. (2020). Artificial Intelligence in Digital Credit in India. Dvara Research. Retrieved from, <https://www.dvara.com/blog/2020/04/13/artificial-intelligence-in-digital-credit-in-india/>

Sinha, A., & Mathews, H. V. (2020). Use of algorithmic techniques for law enforcement: An analysis of scrutability for juridical purposes. 55(23). Retrieved from, <https://www.epw.in/journal/2020/23/special-articles/use-algorithmic-techniques-law-enforcement.html>

Smith, C. A. (2018). *The Colour of Creditworthiness: Debt, Race, and Democracy in the 21st Century*. Baltimore, Maryland: Johns Hopkins University. Retrieved from <https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/60992/FORSTER-SMITH-DISSERTATION-2018.pdf?sequence=1&isAllowed=y>

State of Punjab v Balbir Singh, 3 SCC 299 (1994)

Sundar and Ors v State of Chattisgarh, 7 S.C.C, 547 para. 73 (2011)

Terry, N. (2019). Of Regulating Healthcare AI and Robots. *Yale Journal of Law & Technology*, 21, 18. Retrieved from https://yjolt.org/sites/default/files/21_yale_j.l._tech._special_issue_133.pdf

Terry v Ohio, 392 U.S. 1 (1968)

Travancore Rayons v Union of India, AIR SC 862 (1971)

UN ESCAP. (2019). *Artificial Intelligence in the Delivery of Public Services*. Retrieved from <https://www.unescap.org/sites/default/files/publications/AI%20Report.pdf>

Zee Telefilms v Union of India, AIR, SC 2677 (2005)

Appendix: Examples of Regulatory Tools for AI

Accountability, Oversight, and Redress

- Clear, funded, and appropriate mechanisms for redress.
- Systematic and bottom-up impact assessment of potential harms to civil liberties and human rights.
- Detection, mitigation, and response mechanisms for possible errors as a result of initial training and self-learning.
- In-built audit mechanisms and possibility of verification by an independent third-party.
- Clearly articulated liability structures for situations that involve the use of an AI system.
- Mechanisms for consistent and regular evaluation and review of AI systems, including inclusive and bottom-up mechanisms for tracking impact.
- Communication of changes to AI systems resulting from monitoring and evaluation.
- Capacity-building and awareness of data-driven decision making in courts at national, regional, and district levels.
- Clear framework for working with the private sector, including enabling access to training data held by the private actor, opening up source code, and assigning clear modes of contractual liability.
- Certification schemes and trainings for end users.

Equality, Dignity, and Non-discrimination

- Anti-discrimination standards in compliance with constitutional and international human rights laws.
- Diversity assessment for members of development/implementation team.
- Written standard operating procedures (SOPs) during curation of the data and training of the algorithm.
- Mechanism for incorporation of citizen voices and feedback throughout implementation.
- Framework for assessing disparate impact on specific vulnerable communities.

Safety, Security, and Human Impact

- Impact assessment of all cyber threats to which the AI system could be vulnerable.
 - Risk assessment towards identifying unintended consequences prior to development, including in unpredictable environments.
 - Existing cyber security frameworks at a national level.
-
- Depending on the severity of impact, clear safety controls for a human to override the AI system or reject a prompt, recommendation, or decision by the AI.
-
- Regular security audits, patches etc.
 - Framework for data breach notifications and bug bounty programs.

Privacy and Data Protection

- Compliance with national and global protocols on data protection and governance, including consent principles, control over data use, and restriction of processing, right to erasure, and rectification.
-
- Clear regulatory frameworks for personal and non-personal data in existing data sets.
 - Adoption of necessity, proportionality, and “least intrusive” standards to guide the design, development, and use of AI systems.
 - Built-in mechanisms for notice and consent, with possibility to revoke.
 - Ethical practices in collecting and accessing data for training purposes.
-
- Oversight mechanisms for collection, storage, processing, and use – particularly for real-time and long-term collection and use of data.

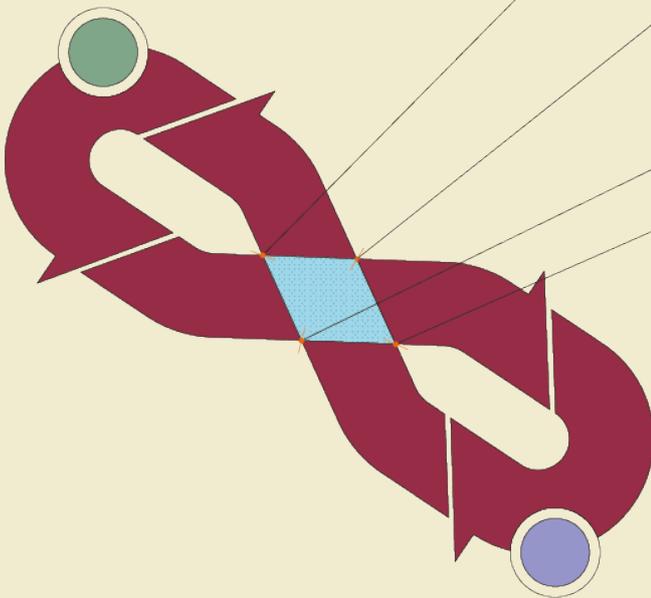
Agency

- Comprehensive notice framework that accounts for passive and active data collection.
- Comprehensive transparency frameworks for data inputs, data training and curation, and use of decisions.
- Retrospective adequation.
- Opt-out options for individuals.
- Gradients of human-in-the-loop.
- Standards for accuracy.

AI Technologies, Information Capacity, and Sustainable South World Trading

Mark Findlay

Centre for AI and Data Governance,
School of Law,
Singapore Management University



This research is supported by the National Research Foundation, Singapore under its Emerging Areas Research Projects (EARP) Funding Initiative. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Singapore.

Abstract

This paper represents a unique research methodology for testing the assumption that AI-assisted information technologies can empower vulnerable economies in trading negotiations. This is a social good outcome, enhanced when it also enables these economies to employ the technology for evaluating more sustainable domestic market protections. The paper is in two parts. The first presents the argument and its underpinning assumption that information asymmetries jeopardize vulnerable economies in trade negotiations and decisions about domestic sustainability. We seek to use AI-assisted information technologies to upend situations where power is the unfair discriminator in trade negotiations because of structural information deficits, and where the outcome of such deficits is the economic disadvantage of vulnerable stakeholders. The research question is the following: How is power dispersal in trade negotiations, and consequent market sustainability, to be achieved by greater information access within the boundaries of resource limitations and data exclusivity? The second section is a summary of the empirical work which pilots a more expansive engagement with trade negotiators and AI developers. The empirical project provides a roadmap for policymakers convinced of the value of the exercise to then adopt the model reflections arising out of the focus groups and translating these into a real-world experience. The research method we propose has three phases, designed to include a diverse set of stakeholders – a scoping exercise, a solution exercise, and a strategic policy exercise. The empirical achievement of this paper is the validation of the proposed methodology through a “shadowing” pilot method. It explains how the representative groups engaged their role plays, and summarizes general findings from the two focus groups conducted.

Analytical Purpose

This paper represents a unique research methodology for testing the assumption that AI-assisted information technologies can empower vulnerable economies in trading negotiations. This is a social good outcome, enhanced when it also enables these economies to employ the technology for evaluating more sustainable domestic market protections.

The paper is in two parts: the initial discursive analysis presents the argument underpinning the assumption; the second section is a summary of the empirical work which pilots a more expansive engagement with trade negotiators and AI providers. This division allows a policy audience to concentrate on the justifications for the assumption, the challenges facing implementation, and the speculated consequences from its successful achievement. Researchers and evaluators will find interest in the details of the pilot methodology.

The paper demonstrates and tests our confidence in the methodology to positively establish the analytical assumptions regarding power dispersal and sustainable domestic market analysis. We advance speculative policy recommendations that can be drawn for the critical experience of the pilot methodology. The paper's commitment to empowerment through policy engagement and recipient ownership makes prescriptive policy inappropriate without a full application of the method in real market decision-making.

Consistent with the overarching project brief, we have identified a need and proposed an AI-assisted answer to that need at theoretical and policy levels. As such, a social deficit is established and a social good through AI is proposed, which is consistent with a major head of the ESCAP development goals. Recognizing resource limitations and time constraints, the empirical project in the second part provides a roadmap for policymakers convinced of the value of the exercise, to then adopt the model reflections arising out of the focus groups and translating these into a real-world experience.

In more detail, the policy and research assumption is that by employing AI-assisted information sourcing, sorting, and analyzing technologies to improve information access and evaluation underpinning economic decision-making, vulnerable economies can better determine sustainable domestic market policy against enhanced trade bargaining capacity. The availability of AI information-assistance technologies (and associated expertise/education)¹ will, it is argued, provide the material and understandings necessary (but currently absent or under-developed) for selecting contexts of domestic market protection to promote sustainability, and for more competently valuing trade bargaining positions in the case of transnational exchange markets.

At a more macro consideration of economic reliance, this policy decision-making enhancement will reduce the reliance on market surplus dumping from more powerful trading partners and its anti-subsistence consequences. As domestic market sustainability is more strategically prioritized, these vulnerable economies will better weather post growth, or de-growth global economic trends.

As for enhanced trading capacity, and specifically empowered trade bargaining positioning, AI information-assistance technologies for data access, automated data management, and analysis, it is argued, will offer social good outcomes to presently disempowered multi-stakeholder trading players who currently negotiate under information deficits and resultant weakened bargaining capacity. AI information-assistance technologies will strengthen bargaining power, which will increase trading revenue and make more achievable aspirations for "world peace through trade" (Dikowitz, 2014).

1. It is not the intention of the paper to specify these technologies. In fact, essential for our belief in recipient "ownership", any eventual policy applications should involve recipient economies in a dialogue with AI technical resource personnel and donor agencies, to determine the technologies best suited to need on a case-by-case basis.

Background

The foundations of our thinking grow from the following propositions, which can be viewed as policy underpinnings:

1. General principles can be identified as governing successful trading bargains;²
2. Trade negotiations usually reflect the relative market power and positioning of participants;
3. Trading partners from more vulnerable economies may require external bargaining support if structural power asymmetries are to be dispersed in their favor;
4. A “free trade model”³ has negative impacts in weaker economies being required to open up their markets and remove protections over domestic social production.⁴ This trade liberalization has meant that domestic market subsistence and economic sustainability are diminished in favor of trading exploitation;
5. Weaker economies have been adversely affected by discriminatory trading arrangements and exclusionist trading alliances, particularly as their trade commodities are undervalued, and their attractiveness as preferred partners is equally so;
6. Automated data management⁵, access to big data⁶, and artificial intelligence technology capabilities⁷, if affordably available to weaker trading economies, offer capacities to strengthen their positioning in certain trading arrangements;
7. A protectionist regression in domestic trade arrangements among major trading powers,

and moves from multi-lateral to bi-lateral trading alliances, both designed to reduce individual trade deficits and to penalize offending trading partners, may offer opportunities for weaker trading economies to assert domestic social production and bi-lateral advantage. The reasoning behind this view is that domestic market liberalization North to South World, ignoring how vulnerable may be the target domestic resource market, leaves vulnerable economies even more exposed to trade discrimination when major global trading nations are reverting to selective and self-interested tariff protectionism;

8. The paradox between free trade open market liberalization, and intellectual property and data transfer protection, disadvantages weaker economies with lower levels of IP “ownership” and effective data transfer controls.

Taking these fundamentals as given⁸, the first part of the paper builds the following argument:

- Employing bargaining theory, a typology of successful trade bargaining can be established and the significant factors, prioritized;
- Anticipating that information deficit regarding key aspects and dimensions of any particular trade bargain will further disadvantage weaker parties,⁹ access to information and critically appreciating its analytical value will level the bargaining power asymmetries;

2. What is meant by “trade bargains” or “trade negotiations” here is specific trade deals rather than prevailing or permanent trade agreements and partnerships.

3. As a policy to eliminate discrimination against imports and exports, the free trading model has never fully been achieved globally. In such an ideal trading frame, buyers and sellers from different economies may voluntarily trade without a government applying tariffs, quotas, subsidies, or prohibitions on goods and services. Free trade is, therefore, proposed as the opposite of trade protectionism or economic isolationism. Instead of freedom and fairness, having attained comparative advantage in production, the hegemon is typically impaired by artificial trade barriers in its quest to penetrate the domestic economies of competing states. Thus, as a state rises from the core to hegemony, it will progressively favor lower tariffs and move towards a free trade doctrine for import receiving markets, while at the same time resorting to tariffs on imports where they are deemed to correct trade imbalances against their benefit. In de Oliver M. (1993) “The Hegemonic Cycle and Free Trade: the US and Mexico” *Political Geography* 2/5: 457-474.

4. Can social production at home be an adequate substitute for market production from producers abroad, particularly when it comes to high-tech commodities and services? The same could be asked about specialist natural resources which are the material life blood of high technology, and as such, trading priorities. We advance here that trade is necessary for balanced development, but trade deals need not crowd out domestic social production through the export dumping of subsidized or cheap replications of sustainable domestic social production.

5. This refers to the application of algorithmic technologies in cataloguing and mapping data at rest and in action, thereby lessening the prospect of “drowning in big data”, <https://erwin.com/blog/automated-data-management-stop-drowning-data/>

6. The term “big data” has come to mean some form of “value-added” data application potentials. Simply, big data refers to extremely large datasets which may be analyzed computationally to reveal patterns, trends, and associations, particularly concerning human behavior and interactions. The size of these sets and their capacity to cross fertilize creates negative challenges to evaluating data sources and their progressive integrity.

7. The paper prefers the definition provided by Stuart Russell and Peter Norvig (2010) *Artificial Intelligence: A Modern Approach* (3rd edition), New Jersey: Prentice Hall; “the designing and building of intelligent agents that receive precepts from the environment and take actions that affect that environment”. This approach connects with a key idea relevant to the present discussion, that AI is not the same as information – it is technology that helps us process information to take actions in the world.

8. It is possible for each of these assumptions to be empirically tested and contextually validated. However, for our initial purposes, they are designed to form the foundations of wider analytical projections.

9. Rather than talking about economies in terms of stages of development, this paper distinguishes participation in economic decision-making and trade bargaining in terms of the relative strength and weakness of participants. Vulnerability is the approach taken here as an empirical measure of relative market power, which can be corrected through more equal access to the information underpinning strategic economic decision-making.

- Understanding the dynamics of a global free-trading model, and its critique in the recent return to protectionism, projections could be offered regarding how weaker trading economies might be advantaged by interventions to improve their individual bargaining power, and at the same time strategically protecting their sustainable domestic social production;
- Information deficits regarding crucial trade bargain variables disadvantage parties¹⁰ with reduced or restricted access to such information;
- Automated data management, access to big data using artificial intelligence technology, and enhanced analytical expertise/education can provide external assistance to disempowered trading parties when seeking to improve their bargaining status;
- Such information access capacity is made more viable through enhanced internet access;
- Aid and development agencies, international organizations, and private philanthropic entities can provide the financial backing to finance the necessary technology for trade information empowerment. Additionally, multi-stakeholder trading arrangements could fund AI information technology capacity to advance aspirations for “world peace through trade”;
- Access to information alone will not rebalance trading power asymmetries. Along with more access, there is a need to invest in critical and resilient analytical capacity.

Each of the paper’s policy underpinnings represent commitments to the greater trading sustainability of small and less powerful trading economies, in a global context where these economies can teach the North World much about sustainability in a post growth, or de-growth trading age. In addition, more encompassing policy eventualities directed to sustainability for vulnerable economies will be enriched by this research through the suggested potentials it offers to enhance informed decision-making about what domestic resources should be retained in domestic markets, and where these market can be opened up to trade without endangering the resilience of such economies.

Part I

The Analytical Challenge

Trade has become essential for the viability of today’s exchange economies, big and small. Global trade that produces benefits for all is also seen as a positive aspect of global governance and peacemaking. Commodities traded will vary, largely depending on the demographics of the economy and its historical development. If we accept that “property is a fundamental social practice” and “ownership is indeterminate” (Humbach, 2017) then there needs to operate a sustainable frame for things traded between parties that want what property and ownership they claim, to work best for their complex social needs.

Unfortunately, as Joseph Stiglitz has observed at the forefront of free trade policy marketing operating from a beggar-thy-neighbor perspective to beggar-thyself (Stiglitz, 2002a), the “free trade” panacea did not realize universal benefits across the globe.

International economic justice requires that the developed countries take action to open themselves up to fair trade and equitable relationships with developing countries without recourse to the bargaining table or attempts to extract concessions for doing so (Stiglitz, 2002b).

Implicit in this recognition of requiring fair trade initiatives driven from the rich and powerful down to the poor and powerless, is pragmatic structural and process cautioning about unequal bargaining relationships. The cynic might say that fair trade is a non-sequitur. A good bargain benefits one to the detriment of the other. If this is the inevitability of trade, at a global level it explains the inequitable and destructive trajectories of contemporary global economic imperialism (Hardt & Negri, 2001). This paper does not proceed within any such inevitability. Nor does the paper ignore that the introduction of AI-assisted information technology can have the

10. Parties to economic decision-making and trade negotiations may be state actors, commercial agents, or multi-participant stakeholders.

unintended adverse consequences of increasing unfairness if the nature of trading biases based on wider hegemonic disempowerment is not appreciated. Laws against protectionism and promoting free trade North to South worlds often give “fairness” a low priority. Along with more access to information, we would encourage the development of legal regimes respectful of, and not simply exploiting, global economic disparity.

When reflecting the problems associated with transferring misunderstood or misconceived concepts of “fairness” into complex socio-technical systems, Xiang and Raji conclude that “fairness” is a mutual enterprise between AI-creators and legal policymakers:

If the goal is for machine learning models to operate effectively within human systems, they must be compatible with human laws. In order for ML researchers to produce impactful work and for the law to accurately reflect technical realities of algorithmic bias, these disparate communities must recognize each other as partners to collaborate with closely and allies to aid in building a shared understanding of algorithmic harms and the appropriate interventions, ensuring that they are compatible with real-world legal systems (Xiang & Raji, 2019).

New Global Economic Models

Sustainable world trade in an era of post growth or de-growth,¹¹ is facing challenges from the push for protectionism and isolationism against trade liberalization and the “wealth of nations”. National self-sufficiency has incrementally been downgraded by free trade imperatives in favor of the internationalization of economic activities. Populist backlash would selectively reverse the forces of global economic engagement in preference for trading imperatives governed by domestic surplus and offshore relative disempowerment.

The potential downsides of free trade are said to be mitigated by:

- Allowing for innovation and structural change;
- Increasing employability and enabling life-long learning; and
- Redistributing globalization gains more-equally in domestic economies through taxation (Reichel, 2018).

Debate these eventualities if you will, but their achievements are no doubt dependent on which side of the globalization engine one sits – is it for prosperity and peace, or alternatively, for intra-country wealth through production chains skewed to stronger economic bargainers?

The political and economic reality of current trade agendas is that vulnerable economies will be negatively impacted via protectionist policies enforced by major trading nations, in different ways but to similarly disabling extents as they were when forced to expose their own markets to the unbalanced influence of North World free trade expansionism. The inequalities of free trade and selective protectionism, operating on profound imbalances in trade capacity, represent the context for policy reform advocated in the remainder of the paper.

Specifically, the policy reform advocated in this analysis involves:

- Recognizing that sustainable global economies will not be advanced by a heavy regression to selective protectionism or a blind adherence to discriminatory and unbalanced trade liberalization.
- Appreciating that free trade can continue as a dimension of positive global engagement where free trade agreements allow for domestic social production and thereby advance the aspiration for world peace through trade.

11. These are several definitions of de-growth which largely focus on economic policy which concentrates less on economic stimulus than sustainable social welfare. For this paper, the concept also incorporates “post-growth” – economic inevitabilities which see growth slowing or flattening irrespective of political and market intervention. See Azam G. (2017) “Growth to De-Growth; a brief history” <https://www.localfutures.org/growth-degrowth-brief-history/>. “[De-growth] challenges both capitalism and socialism, and the political left and right. It questions any civilization that conceives freedom and emancipation as something achieved by tearing oneself away from and dominating nature, and that sacrifices individual and collective autonomy on the altar of unlimited production and the consumption of material wealth. Capitalism has brought further ills such as the expropriation of livelihoods, the submission of labor to the capitalist order and the commodification of nature, (for the South World in particular). This project to establish rational control over the world, humanity and nature is now collapsing.”

- Realizing that the current financial sustainability of vulnerable South World economies, despite those being economies more likely to adjust successfully to post-growth or de-growth regimes,¹² will be enhanced if their bargaining power in trading arrangements, and their capacity to discriminate between what should be traded and what should remain a domestic resource, is empowered through greater information access and analysis.¹³

The next section looks at a model of bargaining dynamics. In particular, it identifies the importance of access to information for empowering bargain participants.

Bargaining Theory¹⁴

What factors determine the outcomes of specific trade negotiations? What are the sources of bargaining power? What strategies can help in improving a party's bargaining power?

Trade bargains can be epitomized as at least two parties engaging for the purpose of some beneficial outcome (which might or might not be mutual) but who have conflicting interests over terms. These common interests are in cooperating for trade; the conflict lies in how to cooperate.

Taking a more contextual approach, understanding the dynamics of bargaining from the perspective of disadvantaged parties in particular, provides an opportunity to appreciate market dynamics and relationships (internal to the bargain) as well as the influence of political and economic policies' repositioning transactions (external). Interrogating the essential features of the bargain requires more than disentangling reasons for agreement or disagreement. A power analysis is at the core of bargaining theory, governing the imperatives for gaining the best benefit, and often at the cost of fairness or other more universal normative considerations.

Practically, issues of efficiency and distribution are important. Efficiency is at risk if the agreement fails or can only be reached after costly compromise and delay. Distribution relates to how gains emerge from co-operation between the two parties. To these issues identified by Muthoo, we would add sustainability. It is rare that trade relationships are "one-off's". They usually lead on to the establishment of enduring market connections, or they have ramifications for the parties involved, which stretch beyond the commercial terms of the deal.

What are the determinants of the bargaining outcome?

A. Impatience, or the pressures of time

Each player values time. The preference is to agree to the price today rather than tomorrow. The value given to time will be subjective and relative. In particular, it may be as disproportionate and incremental as it is exaggerated by other external cost pressures. Weaker players may have less time to bargain or stronger players may exert the pressures of time if the rapid conclusion of the bargain is essential for other bargains to follow.

Apparent impatience can lead to a weakened bargaining posture or a breakdown of other rational communication essentials. In order to avoid the exposure of impatience, bargaining theory suggests that the vulnerable party should decrease their haggling costs and/or increase the haggling costs of the other party. *One way of achieving such a differential is for the otherwise impatient party to possess and understand the richest range of information and data that constructs (or constricts) the other party's bargaining context.*

Because the wealth and power differentials between trading parties are structural (and often not temporal or spatial), a basic principle of bargaining theory is that economies are unlikely to converge in wealth and income solely through international trading policy.

12. Some say that developing economies need the benefits of growth before adopting a largely North World economic countermovement like de-growth. There is an alternative argument that the conditions required for rethinking the place of the economy within the social, and prioritizing social rather than material goods, are more apparent and resilient in less modernized and less materially dependent societies. The debate is usefully discussed in Lang M. (2017) "Degrowth: Unsuitable for the Global South?" *Alternautas*.

<http://www.alternautas.net/blog/2017/7/17/degrowth-unsuitable-for-the-global-south>. In any case, we are not requiring de-growth, but rather post-growth approaches to sustainability that accept growth as a priority for the South World but in the context that economic growth is repositioning as a global economic agenda.

13. In advancing this thesis, we are mindful that information access alone will not empower market stake-holding. The quality of that information (i.e., its relevance, immediacy, and analytical transparency) all depend on more than technological facilitation. The factors on which information empowerment relies are contextually important when evaluating the significance and sustainability of technological facilitation.

14. The following summary draws heavily on Muthoo, A. (2000) "A Non-technical Introduction to Bargaining Theory" *World Economics* 1/2: 145-166

Features integral to bargaining dynamics such as information deficit, we argue, have greater potential to counterbalance prevailing structural inequalities that determine patience to let negotiations run their natural course.

B. Risk of breakdown

If while bargaining, the players perceive that the negotiation might break down into disagreement because of some exogenous and uncontrollable factors, then bargaining dynamics will alter. Risk of breakdown can be raised through a range of variables from human incompatibility, to the intervention of third parties.

This risk perception is where strategies to increase risk aversion are important. *Information available to parties concerning the nature of the risk and its impact on the other side becomes important if a weaker party wants to shield through risk aversion.*

C. Outside options

Here, the principle is that a party's bargaining power will be increased if their outside option is sufficiently attractive – that is where alternative trading/ bargaining arrangements may parallel the first instance bargaining. Weaker parties are often devoid of any other option, outside or otherwise, or because of not fully understanding the values and variables at play in their bargain, feel trapped within a trade that is anything but to their advantage. *The outside option principle is directly impacted by the amount of information either or both parties have about the bargain in play and the outside option relative to the first instance bargain. The valuation of an outside option will depend not only on the conditions and characteristics of that option, but as much or more on its consequences for the bargain in play.*

D. Parties' relationships

There is much in bargaining theory which concerns the significance of connections between the parties in contexts outside the bargain in hand. These externalities (such as cultural familiarity and political bonds) may impress so deeply into every other condition of the bargain, that negotiations cannot break free from responsibilities and obligations inherent within any such prevailing relationship.

Again, information imbalance, or data access restrictions built into such extraneous relationships will further exacerbate the information deficit retarding knowledgeable participation in the eventual agreement struck.

E. Parties' interests and preferencing

Individuals and organizations seeking to influence economic decisions or to achieve success in a trade bargain, approach the enterprise with pre-formed preferences and exhibiting internalized interests. The decisions or bargains with which the result will be colored by such preferences and interests in the same way that any market choice is in part the product of preference gratification, interest, containment, or satisfaction. Pound (Grossman, 1935) would see the contest over interests as settling on individual claims, demands, or desires. How any of these features have a preference through a bargain or decision will reveal the relative power exercised by individual stakeholders, and by dominating any conflict over interests, the way power differential may be increased.

In trade negotiations, the interests of stakeholders will range well beyond the remit of what is to be bargained or decided. *Therefore, if the influence of pre-existing preferences and interests is going to weigh significantly on the negotiation or decision-making dynamics, then the more each party has detailed and informed knowledge about these preferences and interests, the less likely they will distort outcomes in ways which could not be planned for or at least anticipated by negotiating parties on both sides.*

F. Commitment tactics

In many bargaining situations, the players often take actions prior to/or during the negotiation process which partially commit them to some strategically chosen bargaining positions. If these commitments are partial in that they are revocable, depending on how far down the line of negotiation they have been struck, this may progress the appearance of intractability and therefore costs associated with their revocation. Many of these commitments may have been orchestrated in order to increase the “bluff” (e.g., the limitations on a party to negotiate freely beyond the terms of another commitment). *The power of bluff is always dependent on contrary information or any suspension of disbelief in the bluff.*

G. Asymmetric information

It might be accepted bargaining practice that one party will always know something the other does not. How such an information disparity should be valued is relative to the significance of the information for the vital terms of agreement (or disagreement). Information asymmetries affect both the values and pricing on offer as conditions of a deal, as well as when the agreement might be concluded for the maximum mutual benefit.

In general, an absence of complete information will lead to inefficient bargaining outcomes, even for those who benefit from an information surplus. The logic behind this view rests on an acceptance that the more information available to both parties, the earlier synergies will be established and bargains struck.

The message is that treating information in some exclusionist or proprietorial manner may produce a short-term bargain benefit for the information owners (renters and possessors), but at the risk of an unsustainable trading market vulnerable to misrepresentation, exploitation, corruption, and the retarding on any natural propensity for market competition.

Therefore, policies to defeat information asymmetries in trading arrangements, we argue, offer empowerment potentials for weaker players, and on the strength of power dispersal through information access and sharing, more sustainable trading markets ongoing.

In seeking power dispersal via information access and analysis, this paper is not requiring some egalitarian levelling of market engagement. As Rawls argued, social inequality will not always be the product of power abuse or discrimination (Grcic, 2007). What we are seeking to attack are those situations where power is the unfair discriminator because of structural information deficits, and economic abuse of vulnerable stakeholders is the outcome.

From this review of bargaining dynamics, the essential research question emerges: How is power dispersal in trade negotiations, and consequent market sustainability, to be achieved by greater information access within the boundaries of resource limitations and data exclusivity?

Access through AI, Automated Data Management, and Big Data – Some Critical Considerations

As suggested in the brief reflection on fairness (above), it is necessary to preface any consideration of the relationship between improved data access, and improved bargaining power in trading arrangements, with the caution that more data and better automated data management courtesy of AI technologies will not automatically empower weaker trading partners. In fact, increased technological capacity to access data, unconnected with significant advances in data appreciation and contextualization may simply further fog the understanding of smaller stakeholders and exacerbate bargaining disempowerment.¹⁵

In addition, bargaining tactics may prefer privacy when information is applied, sought, withheld, or

15. The focus group discussions in the Methodology section enunciate this concern.

exchanged. The bargaining attitude that bargaining power is lessened if information is mutualized has to be addressed with the argument that for market sustainability, and not just a single bargain advantage, fairer information access will make for more robust economic engagement. Once again, we return to the externalities of economic fairness.

How market stakeholders accommodate and benefit from information abundance is at the heart of any policy derivatives designed to improve trading balance in a hegemonic global trading model, intensified in its potential to discriminate as a consequence of selective and politicized protectionism. A feature of the methodology to follow is the potential to better understand how information needs to be met with enhanced information access to address specific bargaining decisions.

An important consideration, which informs the policy projection for trade empowerment, is its timeliness. With the major trading partners at war over tariffs, trade imbalances, protectionism, and perversely, secrecy when it comes to tech transfer and IP, the conditions may be right for smaller trading economies to rebalance their domestic sustainability without the backlash of free trade essentialism.¹⁶ From that stance, an informed and economic evaluation of what remains open for trading will provide a more stable platform for trade bargaining.

Access to information, complemented by increased analytical capacity, will enable more nuanced distinctions between protection for domestic sustainability and competitive positioning in regional and international trading. Yet, strategic analytical capacity does not simply depend on more devices and bigger technologies. In fact, the savvier information-users are navigating away from an over-reliance on devices and are becoming aware about how algorithms affect their lives. In any case, even those market players who have less information are relying

more on algorithms to guide their decisions, whether they realize it or not. In the current technologized world environment, it is axiomatic that new digital literacy is not about more skillfully using a computer or being on the Internet on call, but understanding and evaluating the consequences of an always-plugged-in lifestyle for every aspect of social and economic engagement. In societies and cultures that still place social relations much above digital connections, the introduction of AI capacity is never, as we see it, meant to diminish or downplay the dominant role of human agency.

Over two thirds of the worlds' population either live outside or can only partially participate in the digital age. Digital access and digital literacy are now recognized as fundamental human rights. However, when it comes to fair trading practice, a level playing field in terms of information engagement is not only a long way off, but some might argue is a misunderstanding of bargaining behavior and advantage (UNCTAD, 2019).

In 2014, the UN General Assembly adopted resolution 69/204 "Information and Communication Technologies for Development". Most relevant for this paper is the reference to:

"... information and communications technologies have the potential to provide new solutions to development challenges, particularly in the context of globalization, and can foster sustained, inclusive and equitable economic growth and sustainable development, competitiveness, access to information and knowledge, poverty eradication and social inclusion that will help to expedite the integration of all countries, especially developing countries, in particular the least developed countries, into the global economy" (UNCTAD, 2015);

This paper is not solely concerned about a "digital divide" between those who have access to computers and the Internet and those who do not. As digital devices proliferate, the divide is not just about access

16. As noted earlier, there has been much political hypocrisy surrounding the "freedom" of free trade, and as such, a re-balancing of domestic sustainability and regional/international competitiveness will not necessarily require a wholesale rejection of more open cross border commercial engagement.

17. The mirror image of this divide is the incapacity of algorithm designers to appreciate the complexity and sometimes intentional ubiquity in the social circumstances and human decisions to which they are applied.

or available technologies. How individuals and organizations deal with information overload and the plethora of algorithmic decisions that permeate every aspect of their lives is an even more relevant discriminator when turning a power analysis to the global trade divide (Susaria, 2019). The new digital divide is wedged over understanding how algorithms can and should guide decision-making.¹⁷

*The “empowerment through data access and analysis” model that is advocated here depends on the availability of technological facilitation in identifying relevant data, determining its legitimacy and fitness-for-purpose, alongside enhanced analytical capacity and an upgraded appreciation of how AI as information technology can enhance essential economic and trade decision-making.*¹⁸ Along with this external impetus for empowerment in decision-making is a concurrent challenge for information users in vulnerable economies to more clearly determine who decides what technologies should be preferred and whether such technologies offer decision-making options that are fair/legitimate/fit-for-purpose.

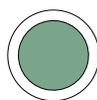
Syncing AI potentials with the information needed for domestic and trans-national trade bargaining and economic decision-making, is not singularly a question of sourcing and supplying technological capacity presently unavailable to weaker market stakeholders. Along with improved access and analytical technologies, there is a need to target the utility of such technologies and the information they produce to domestic economic sustainability (through trade protection) and increased trading profitability (through sharper trans-national bargaining).

In identifying the necessity for a more level playing field over data access and analysis in trade negotiations, this paper is not traversing discussions of “data trade” and its regulation, nor are we focusing on data driven economies.¹⁹ The policy product of the research to follow is also not seeking to challenge even the most discriminatory IP and data protection regimes, though such challenges might successfully advance market sustainability in an era of access revolution (Findlay, 2017). Rather, the purpose of the research method to follow is to scope the type of information necessary for successful domestic market discrimination and trade negotiations, and the manner in which the provision of access and analytics technology (via AI potentials) can enhance the decision-making benefits which sustainable domestic market analysis and invigorated trading negotiations offer for empowering and assisting vulnerable economies at a time of world trade transition.²⁰

Bargaining-empowerment Through Technologized Information Access and Analysis

Bargaining-empowerment through information access may occur in several ways. Recognizing there is a difference between:

1. access to information helping an individual actor to bargain better, and
2. access to information assisting this actor to locate other stakeholder participants, and together they bargain better (because they share information and they act as a more influential bargaining unit);



18. In identifying this decision-making “space”, we recognize the importance of determining how to increase domestic market sustainability, while at the same time evaluating what should be traded beyond the domestic market and at what value.

19. For an interesting discussion of these two themes and their intersection, see Ciuriak D. (2018) “Digital Trade: Is data treaty-ready” *CIGI Papers No.162* <https://www.cigionline.org/sites/default/files/documents/Paper%20no.162web.pdf>

20. In talking of trading arrangements in terms of state-to-state dialogue, we are, for the purposes of this research, simplifying the trading demographics wherein private sector players may be as significant or more so when vulnerable stakeholders in trade negotiations expose their domestic markets and resources to the interests of external multi-national traders. This paper was settled prior to the impact of the COVID-19 pandemic on global economic relations and as such cannot take these influences into account for this analysis.

21. We recognize that these quality-control problems are exacerbated the bigger and more interconnected are the datasets.

Once information has been identified, its sources need to be understood and the prudential pathways through which it has passed if relevance and reliability are to be measured.²¹ The quality of information matters in terms of its decision-making value, and information offers diminishing decision-making returns as that quality is less open to testing and verification. A small amount of high-quality information is likely more useful than an abundance of low-quality information. In that regard, access and analysis must be accompanied by easy methods for data evaluation against simple matrices. An example of the variables to be considered would be (where visible) completeness, timeliness, uniqueness, accuracy, validity, and consistency (IT Pro team, 2020).

Next comes the issue of information over-load. Unleashing masses of information, high quality or not, will swamp vulnerable users without the capacity to process it. The other side of this problem is where AI and data analysis technologies can respond for social good.

Is this assertion confirmed by the literature? The studies associated with improved bargaining power as a consequence of greater information analysis are heavily concentrated on labor mobilization.²² Analogies are usefully drawn from this literature insofar as it has a distinct interest in negotiations, bargaining, and decision-making modelling.

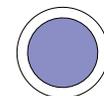
It is not novel to suggest that AI technologies can enable better trading outcomes for vulnerable economies. The United Nations Conference on Trade and Development (UNCTAD) recently introduced a new AI tool to speed up trading negotiations by simplifying complexity. As part of the Intelligent Tech and Trading Initiative²³, UNCTAD and the International Chamber of Commerce have produced a prototype of what they call the Cognitive Trade Advisor.

“Developing countries and least developed countries have limited resources to prepare for trade negotiations,” said Pamela Coke-Hamilton, Director of International Trade and Commodities of UNCTAD.

“The amount of information that negotiators and their teams need to process is proliferating, and often they need the information on a timely and rapid basis,” she said. The Cognitive Trade Advisor uses an understanding of natural language to provide cognitive solutions to improve the way delegates prepare for and carry out their negotiations.

“The texts of the agreements are getting longer and longer,” Ms Coke-Hamilton said. “In the 1950s, an average trade agreement was around 5,000 words long. In the current decade, this has increased to more than 50,000 words. Dealing with such amounts of information takes a lot of time (UNCTAD, 2018).”

Interesting as this development might be, our policy frame has a more restricted but no less impactful intention. As mentioned earlier, we are not touching on preferential trading arrangements, or the understanding of their complex documentation.²⁴ Instead, our remit is more contained, and as such, attainable without new technologies. The direction of the policy to follow is the employment of presently available AI technologies for accessing and analyzing information that can better position vulnerable negotiators by reducing crucial information deficits. The UNCTAD initiative is to develop new AI tools in order to make the attainment of Sustainable Development Goals more likely in under-developed regions. This paper shares the desire to see AI supporting progress to these goals by reducing negotiating inequalities. On the way to achieving this aim, the poverty in AI experience with currently available technologies in vulnerable markets and societies will hamper developments towards these goals even before new, affordable, user-friendly, and sustainable technologies are more readily available.



22. An example is <https://turkopticon.ucsd.edu/>

23. Information retrieved from <https://itti-global.org>

24. For example, see Alschner, W., Seiermann J., & Skougarevskiy, D. (2017) "Text-as-data analysis of preferential trade agreements: Mapping the PTA landscape" UNCTAD Research Paper No. 5. <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=1838>

Employing AI-assisted technologies for information access may or may not be in itself a neutral endeavor. In advocating this progress, there needs to be sensitivity to political and cultural parameters in offering AI technologies to analyze and prioritize economic and trading decision-making. Many post-colonial vulnerable economies do not respond well to top-down capacity building from the North World, especially when North/South disempowerment is identified in these economies as the root cause of their trade problems in the first place.

It is not the intention of this paper to provide a pre-packaged menu of preferred technological options to enhance access and analysis. As the methodology section to follow sets out, "ownership" of this selection should be offered through a scoping exercise which identifies context-specific needs and solutions. Ultimately, the preferred technology should be seen by the potential user as at base beneficial and manageable within the specific dynamics of their decision-making and bargaining ecology.

In seeking to identify the types of AI-assisted information technology that would best support vulnerable economies in domestic resource economic decision-making and trade negotiations, the following factors are important selection criteria and *determine how the policy suggestions in this analysis should be implemented*:

- *The technology needs to be affordable.* Even if its purchase is subsidized there are running and maintenance costs which will fall to the user and as such, these need at least to be defrayed by cost-savings through improved decision-making outcomes and bargain positioning.
- *It must be user-friendly* and explicable so that institutional, cultural, or administrative resistance to new technologies, or suspicions about the hidden agendas they might translate from donors, can be overcome.



- *The technology must be robust and resilient.* The anticipated user population will not be sufficiently resourced with sophisticated tech support to manage frequent and constant hardware and software upgrading.
- *It should be capable of timely employment* in the various vital stages of decision-making and bargaining.
- It must have *rapid analytical capacities.*
- Its operational language must be *in sync with the language of the bargainers and decision-makers.*
- On the basis of the information it accesses and analyses, *it should provide cognitive solutions from which the participants can draw informed choices.*

On the nature of information absent for access by policymakers looking at trade and domestic resource market sustainability from the perspective of vulnerable emerging economies, the imperial influence of platform distributors over raw data is an important reflection in the empowerment equation.

The commercialization and monetarized analysis of raw data through the big platforms presents a significant challenge when approaching the issue of more open access as an empowerment policy (UNCTAD, 2019). Accepting that there will always be sensitive metadata driving information technologies and linking through even simple keyword searching to an array of mediations over raw data for commercial purposes. The present project cannot neutralize this phenomenon, but it can flag it as a further level of potential disempowerment and seek transparency and explainability of data sourcing and technological translation in a language that the end user can appreciate and take into account when relying on information.

It is with this caution in mind that the project methodology is advanced.

Part II

Methodology

The project's methodology involves a pilot stage, the results of which are summarized in the conclusion of this section. Having satisfied ourselves that the focus group methodology is appropriate to test the analytical underpinnings, the project-proper methodology is described for later implementation.

The methodology has two clear underpinnings. First, to adopt a top-down approach to empowerment, with stakeholders already distrustful of the motivations which may underlie the actions of parties who in the past have been seen as complicit in the disempowerment reality, would endanger the sustainability of the support provided. Aligned with this is the second concern that both the research and the policy outcomes it supports should form stages in the empowerment process.

Therefore, the initial context for designing the first component of a research plan is to appreciate the nature of decision-making vulnerability that trading policy will need to address, and sustainability evaluations will need to constantly be monitored. Vulnerability is not to be viewed only in terms of power imbalance, or to substitute for terms such as "weakness", "disadvantage", and "discrimination" (Fineman, 2019).

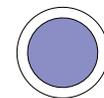
Applying this individualist conceptualization of vulnerability to economies, markets, and societies, we can imagine a research method that appreciates the forces which create and maintain vulnerability, and provides a voice to the disempowered that resultant policy is designed to enable. In particular, and working from our earlier review of bargaining theory, the research should test whether decision-makers from

vulnerable economies realize information deficit in terms of decision-making need, can articulate the sources and substance of information that would be useful to them, and from there speculate on how such information wants to be analyzed, validated, and sustained. Once this "needs analysis" has been trialed, it then becomes the task for information technologists, with an understanding of information disadvantage and its decision-making context, to suggest AI-assisted information options that could empower sustainable decision-making.

The research design in its post-pilot phase has three phases:

Participant Focus Groups

In the format of a facilitated focus group, a series of hypotheticals designed to provoke situations of vulnerability in economic decision-making and trade bargaining will be put to a meeting of negotiators and policymakers who have experienced disempowerment through information deficit. Recognizing the risk of "digital imperialism" in designing research experience from an external AI focused context, these hypotheticals will have been previously discussed, critiqued, and settled by a small working party drawn from scholars, negotiators, and policy people with a familiarity of South World economic disempowerment. In particular, the advice on drafting the hypotheticals will be taken in the first instance from experts in mediation and negotiation with hands-on experience of South World decision-making styles. Added to this will be the impressions from policymakers and negotiators working in South World trade and economic environments.



a) Policymakers Focus Group, Scoping Exercise

Participants in this focus group will be drawn from five nominated economies with currently unsustainable domestic resources, limited trading advantage, and who it might be argued, have suffered as a consequence of North/South World market liberalization.²⁵ The output from this focus group would be a clear understanding of the information needs of participants, the situations in which the absence of specific information on which to rest decisions or make bargains is deemed to disempower, suggestions concerning what information access needs to be prioritized, and the types of cognitive options that participants would think helpful and why.

b) Expert Focus Group, Solution Exercise

Armed with the information disadvantages identified in the first focus group, experts in the field of AI technology, development capacity building, trade negotiation and economic decision-making, mediation, and multi-participant stakeholder policy work would be charged to apply particular AI technologies to the problems represented in the first set of hypotheticals. In addition, members of focus group 2 will have access to a structured transcript of the discussions emerging out of focus group 1. Using the same hypotheticals across both focus groups will offer some qualitative consistency and comparability. Participants in the first focus group could be invited to attend and observe these discussions. The output from this focus group would be the preparation of a set of AI technology options nominated against the particular information deficits identified by the first focus groups. In preparing and costing these options, participants would be asked to reflect on the list of selection criteria that is described above.

c) Implementation Focus Group, Strategic Policy Exercise

The final focus group would involve academic experts in vulnerability and social justice, as well as negotiation/mediation, policy regulation, and social

development, similar to those drawn together to formulate and test the hypotheticals. The Emory/Leeds Vulnerability Initiative and scholars with interests in law and development, negotiation and mediation, and information systems would facilitate a policy forum designed to produce a workable policy agenda for information empowerment and market sustainability in the five nominated vulnerable economies. Additionally, experts from global information and communication organizations with abilities to fund a pilot scheme, representatives from ESCAP with responsibilities for promoting the UN Sustainable Development Goals, and interested participants from the previous two focus groups, would add to the policymaking dynamics. The policy yield from this workshop would be to roll out a pilot program that would enable an empirical evaluation of the impact of AI technology capacity building on the achievement of better trade bargaining benefits, and sustainable economic decisions regarding the safeguarding of domestic resources in vulnerable economies.

“Shadowing” Pilot Focus Group Method-validation Exercise²⁶

“Shadowing” is a style of simulation, where the survey population is brought together (usually at a pilot stage) to represent the intended actual survey population for the purposes of testing whether the research methodology is promising and potentially reliable. Shadow methodology is where the survey population is asked to assume the roles and responsibilities of an actual population, and where possible, to follow the progress of that population as it performs a particular decision task. This method has a history in jury research, in the US.

For the purposes of piloting, a combination simulation/shadow methodology was applied through two focus groups. The first identified information deficiencies in trade bargaining and domestic resource sustainability among trade and development policy personnel. The

25. APRU member institutions and their affiliates will be helpful in identifying and facilitating participants.

26. Due to pressures of time and limited resources, the pilot was not able to target policy makers in vulnerable economies (focus group 1). However, it was possible to engage with young AI technical experts in focus group 2. The implementation focus group 3 was not necessary at the pilot stage.

second group, with the benefit of such identification, then offered AI-assisted technology options. Hypotheticals enabled information and decision-making simulation to be experienced and monitored, and assembling a group of participants who were instructed “in-character” provided the “shadow” survey population with capacity to identify information disadvantage prompted by the hypotheticals.

For both groups, a small number of participants with similar social demographics²⁷ were drawn together.²⁸ The first group was asked to take on the character of a trade negotiator, or a trade and development policy officer in a nominated emerging economy.²⁹ Along with the identified role and jurisdiction, each participant was assigned a particular strategic concern in performing their function. That concern was connected back to limitations in the information base, or information deficits effecting the potential of each player to make knowledgeable trade policy or bargaining decisions, and determinations about domestic resource market sustainability.

From that perspective, and prior to focus group 1, each participant was encouraged to research their role with limited direction from the focus group administrators. The reason that this research stage is unstructured relates to the expectation that members of the actual population (trade policymakers, trade bargainers, and domestic resource decision-makers) will possess different levels of knowledge and experience depending on personal, professional, and information-centered variables, as well as differing degrees of self-reflection.³⁰ The only direction given for this independent, individual research phase was the necessity to focus on what information sources, technologies, and analytical capacities exist in the nominated jurisdiction. Even the actual population would be differentially challenged to know and identify what information is missing and what is needed due no doubt to varied personal experience and confronting variables (structural and functional) that are sometimes difficult to enunciate.

The second focus group was drawn from participants with special skills in AI-assisted information technology options. This group, while not specifically researching the information disadvantages existing in the 5 vulnerable economies, were required to reflect on these in a general sense and were assisted by the transcript from the first focus group, as a discussion and reflection resource.

The hypotheticals designed for group 1 to elicit information deficit and information need, contextualized knowledgeable decision-making and empowered bargaining/resource-retention determinations in three specific directions: natural resources trading, consortium-sponsored foreign direct investment, cash cropping diversification and regional security (see Appendix 1).³¹ With the identification of information/analysis need, focus group 2 were asked to suggest and design practical options from available and affordable AI-assisted technologies directed to trade bargaining, and trade/ domestic resource balance. Sustainability for these options is a priority.

Simulation/shadow survey populations are a compromise at the pilot stage but, with the participants applying sufficient dedication and immersion to their role-play, the discussions unfolded as a useful test-pad for whether this method should be applied in the more resource-demanding environment of actual survey populations. In particular, we wanted to explore whether participants can see the issues with what should be available to them, what they do not know, what is hidden, what-leads-to what in any information chain, and once more information is available, how it can empower the decision-making/ analytical challenge. It proved possible to elicit responses along these lines in the context of the hypotheticals (see General Findings). It also emerged possible from group 1 to group 2 that information deficit, once identified, was followed by technological enhancement, which will lead to more empowered bargaining/decision-making capacity and outcomes.

27. A group of young, tertiary educated men and women with varied knowledge of the essential population experience, but briefed to take on a character within a defined context.

28. Once the participants for focus groups 1 and 2 were identified, they were separately briefed as to the purpose of the shadow simulation and were assigned characters and tasks to research and adopt.

29. The economies selected were Papua New Guinea, the Philippines, Vietnam, Cambodia, and Myanmar.

30. The need for self-reflection is a central tension in the exercise of any focus group method.

31. The testing of hypothetical utility would be another feature of the focus group experience and ownership.

Focus Group 1 – General Findings

Starting out with the MNC/natural resource trading scenario, the initial information need centered on sufficient knowledge about the bargaining partner and the possibility of developing a relationship of trust. In addition to what might be found on the commercial public record, it was suggested to use already existing public and private sector trading networks, and explore previous case-study instances of the operation of the MNC in the region with similar trading conditions. Reservations were expressed about asking the MNC directly, based on different interpretations of power imbalance.

Next, it was deemed necessary to identify major decision sites and bargaining points in the commercial supply chain if the deal progressed. It was noted that some information along the chain might be protected as commercial knowledge. Mention was made about information access costs (material and representational) in contacting third parties and seeking commercial data. Would there be available historical aggregated data on harvesting, processing, marketing, and consumption and where and how could it be accessed? There seemed limited possibilities across other government and private sector agencies in each economy as the natural resource in question was yet to be commercially exploited. International organizations may have relevant data, but because each economy was not already linked to the international standardization networks for this resource, this information might not be easy to access.

Recognizing that this bargain had to be considered against competitive offers (or even exploitation by the state itself) how could other potential investors be approached without damaging the confidence of the deal on the table? What information would be necessary to identify markets for the natural resource, possible market prices, and features of alternative deals that should be anticipated?

Much of this information could come from the MNC itself, but commercial confidentiality may limit this as a source. In any case, information from a trade bargainer would require third party validation. How might this be achieved?

Several participants wanted to ensure that any such trade deal should be the first stage in a commercially sustainable arrangement. Aligned with this concern was interest in the sustainability of the natural resource, and the impacts on a pre-existing subsistence economy relying on the natural resource. Without any detailed natural resource surveys or environmental impact evaluation capacity, what were external analysis options to fill these information deficits?

Assuming that information regarding the MNC, the supply chain, and resource sustainability are available, the group discussed other information needs that impacted on relative bargaining power. It appeared that previous experience in natural resource trading was an important viable. Furthermore, general prevailing trade policy impediments such as nationalized industrial development, institutional corruption, exchange rate interference, and weak trade positioning against major trading economies were identified.

There was discussion about necessary bargaining conditions before negotiations could be progressed besides those already identified. A framework for economic growth and social benefits was identified, but the necessary information on how it might be formulated was uncertain. It seemed clear that in order to see the bargain as having long-term benefit, confidence had to be developed in the MNC's commercial intentions; and again, that would depend on knowledge about the breadth and depth of the MNC's commercial intentions. One participant specified the importance of tech transfer as part of the deal, and the development of feedback loops so that information deficit ongoing would not simply exacerbate misunderstanding and mistrust.

Looking at the consortium scenario, the problem of power imbalances through compounded interests was a recurrent theme. A sense emerged that the bargaining interests of different players were more than they seemed, and how to reveal these was a central information question. Participants wanted to know more about what they were not being told. The experience of other states in dealing with consortium members was suggested as a data source, but problems with confidentiality agreements would arise. In particular, information about the bank's standing within the international financial sector, along with more detail about the terms of the loan and penalties for default were required. The issue of imported labor for the construction company was not considered acceptable because it was not explained beyond skills, and if local labor was not involved there would be no instructional benefit through the exercise. The immigration law implications would lead to a need for "whole of government" information sharing.

The precluded options for power development provoked a need for much more data about the proposed nuclear option, as well as its risks and benefits. Additionally, the rejection of solar options would not be acceptable without some comparative market/environmental analysis.

Particularly, when it came to the push for 5G technology, participants felt totally disadvantaged by knowledge deficit concerning the technology and the implications of coincidental obsolescence. As there was no indication by the consortium of the sustainability of this new technology following its introduction, data about which only the consortium could furnish, participants expressed no position for evaluating cost/benefit. Local business concerns needed development so that they could be put to the consortium proposer for its response, which then would require external evaluation.

When invited to dissect the consortium offer, the fear was expressed that to cherry-pick might mean the loss of desperately needed foreign direct investment. Without environmental impact evaluation for the medium and long term, it was difficult to assess whether the costs attendant on the FDI would outweigh the boost to foreign capital, particularly minus clear capacity building concessions.

The final hypothetical canvassing cash crop diversification also presented regional relations issues. The question of crop security was not addressed and needed to be. However, as with most of the information deficit pertaining to this proposal, there would be a disempowering reliance data sourced from the other bargaining party. This situation emphasized a perennial concern about data validation.

As the arrangement could degenerate into little more than the participant states providing the "farm" for all the offshore commercial benefits, there needed to be information on plans for sustainability, and benefits for the domestic economy. This scenario presented tensions between macro and micro policy desires (diversified cash cropping vs. domestic security and reputational issues), and information was needed in the form of projections on the wider socio-political consequences going forward. Special mention was made about the importance of labor-force benefits, not just through the proposed (but unspecified) R&D injection, but more generally regarding associated agricultural labor transition and mobility. For instance, what would be the concentration on planting/harvesting technology? Participants felt empowered at least to require a detailed business plan from the proposer. Worries about the development of an underground economy in parallel and the exacerbation of already-existing drug problems required thinking through.

Focus Group 2 – General Observations

At the outset, there was general comment that reflected the concerns of those in the first focus group without then often moving to advance specific information technology solutions. This reluctance could have been a consequence of insufficient clarity that we were not looking just to throw tech at any information deficit. In addition, it reflected the group's belief that data and associated information technology gains its relevance from the questions first asked about need.

The most significant takeaway for the facilitator was the need for a two-pronged approach to information disempowerment, which marries mundane data collection and access devices/routines with capacity enhancement among those who will apply the information to the decision-task. This does not come, originate, or exist as a generalized application, and instead requires purpose of design, modification, and infrastructure support, which we did not get to specify in every hypothetical area of need identified in the first focus group. The main impression for the rapporteur was regular reference to needing to know what the data problem was, which required a data resolution: meaning that both the initial need and whatever data collected may satisfy it, should be clearly specified. Obviously, these observations return to a knowledge gap issue and the requirement for capacity building rather than information tech on its own.

An important qualification about the information empowerment thesis is its present over-emphasis in the current project design on state capacity building. Participants mentioned the not-uncommon situation where a state can use information enhancement for purposes which may advance economic interests at social cost. There was also discussion of the need to ensure information empowerment to the private sector, where trade bargaining and resource retention are matters shared between the state and commerce. Finally, in order that the use of data for trading and resource retention decision-making is for social good, any information enhancement project should not leave

out civil society if it is to have the capacity of keeping the other two market players accountable.

Following on from identifying need and sourcing data, discussions included validation and evaluation approaches. With diversity in sources of data, how does one deal with bias? Questions were raised about maintaining the currency and value of data. Original difficulties with knowing what questions to ask might translate into not knowing in what format to employ, store, and order data, or even what the data can accomplish. Added to these are problems of granularity, and the potentially high costs of storage and analysis systems. Connected were worries about giving more data back to companies through information loops and thereby entrenching the information asymmetries in bargaining relativity even further.

On building tech capacity domestically, the data market in the hypotheticals is situated now around identifiable information management needs, so perhaps we are moving into a world where start-ups can be generated without too much capacity required, and these innovators could contribute home-grown information enhancement technologies. How hard would that be? Is it possible to seed something like this? The simplest sustainable solution for information enhancement is to raise capacity within these vulnerable economies to create purpose-designed tech solutions.

On the standardization of data collection vs. having a problem to resolve and then standardizing data afterwards, some participants emphasized "big is best" – the more data you have the better the standardization will be, as well as its application to progressive information needs. A basic observation was made about the utility of producing mundane data from documenting various stages of the supply chain/trade decision-making/auditing processes. Being involved in data production internal to the decision process enables participants to feel that they own that data and understand it better.

Policy Reflections from the Focus Group Deliberations

Information asymmetries on which the project is based:

- Relationship trade bargainer with external partners
 - Necessary to have knowledge about possible trading partners consequential of any trade negotiation – building contacts with such contacts
 - Knowledge about external companies that are offering trade relationships
- Domestic market information gaps
 - Knowledge about demographics of certain markets: e.g., different fishing practices, how fish stocks may be implicated by trade negotiations
 - Knowledge of existing needs of businesses and commercial relationships with or without trade bargain
- Knowledge gaps in technology
 - Emerging technology: target vulnerable economies may be hampered by limited information about these technologies including about servicing and maintenance in the long run. In turn, this traps technology recipients into a relationship with an external organization in the long run which might compromise the sustainability of the suggested tech aid and increase information dependencies

Capacity building considerations regarding asymmetries and dependencies:

- Capacity building to address knowledge gaps in technology and to enable maintenance of technology in the long-run, or to shift away from an over-reliance from single service providers
 - Work to address dependencies concerning data sources, data integrity, and the accountability of tech development. If people do not know what kinds of questions to ask, it will have consequences for data collection, cleaning, processing, and AI-products chosen and employed. In addition,

haphazard or careless data collection may entrench information asymmetries with external data collectors even further and lead to greater inequalities

- A working knowledge of technology would aid clarification of when not to use technology
- For trade negotiators (and the wider associated organizations) working in targeted vulnerable economies that still have limited digitalization and technological capacity, consider steps that would make the collection of mundane data more efficient (and not technology dependent) in the near or mid-term.

Before the injection of AI-assisted information technology

- At the initiation of the project, an intensive needs analysis must be commenced, which is grounded in developing skills around what questions to ask about information deficit, that then will translate into learning about what format to store and order data, and what data can accomplish in trading negotiations and domestic market sustainability.
- Capacity building within the target vulnerable economies will help the identification of major decision sites and bargaining points in the entire supply chain so that negotiators will see where information deficit needs to be addressed.
- International organizations can assist in capacity building as they do not have commitments to either side of any trade bargain. However, due to the lack of relationships between the target vulnerable economy and IOs, consequent on the absence of commercial trading markets on which they may have advised, as well as failure by the target economy in the past to implement international standards, these relationships may need to be project-specific.
- Associated with assistance from international organizations, the target vulnerable economy needs to have access to knowledge in the public domain about natural resources and the demographics of different harvesting practices, and how the relative sustainability of natural resource stocks is impacted

by trading and domestic market decisions. This information access could be provided through aid agencies connection with national scientific repositories and regional data bases.

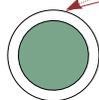
- Target economies must be trained in all areas of government information retention, usage, and exchange, rather than operate with information locked in certain ministries.
- While information technologies are a priority for advanced consumerist economies, this is not the experience in target vulnerable economies. Therefore, prior to the roll-out of such technologies, sponsors and providers should supplement local limited information about servicing a technology, and the dangers inherent in locking into a provider/client relationship in the long-run.
- Along with technological “needs and potential” training, target vulnerable economies and countries in their region will have limited basic market information to do cost-benefit analysis. However, these economies can supplement this information if provided by aid agencies and IOs with essential local knowledge of social/political contexts in which information is best contextualized.
- Through any phase of externally supported capacity building, there is a need to ensure civil society remains in the loop – to understand business needs on the ground.

At the introduction of AI-assisted information technology, and following

- Product sustainability is essential and takes certain crucial forms that must be ensured: data sources – who is collecting data and originally for what purposes; data integrity and validation – how is information to be accredited and verified ongoing?; accountability – ensuring that civil society is

informed about the type of information that is being collected and provided to governments (particularly important when local farms and fisheries are part of the data production chain); technical sustainability of a technical product – who maintains it? These issues require allied services from sponsors, providers, advisers, and locally trained experts.

- Mission creep: if we want to avoid the monetization of technical applications, developers need a clear and disciplined purpose which is struck in agreement with the local end users.
- At the time of introduction, there should be stimulated public debate about intentions around information access and use. Civil society will then be involved in a holistic and integrated approach to data empowerment.
- As a condition of the technology contract, a home-grown sector development and training in technology development in country must be offered. This could be coordinated and stimulated by a developer-centric branch of government.
- Recognize the importance and resourcing of internet penetration into social networks within the economy, particularly those that provide rich sources of resource and market data.
- Recognize the necessity for introduced technology to be affordable, maintainable, and anti-obsolescent.
- Efforts at standardization in the collection of data from multiple sources, which may then enable more actions on the data to be taken in the analysis phase. This endeavor will include leveraging existing collection methods active and accessible prior to the information roll-out. Identification of mundane data built into the consciousness of people who are currently promoting trade and measuring markets (internal and external to the economy).



Concluding Reflections

The rhetoric of what AI can and cannot do continues to be shrouded in mysticism, technological elitism, post-colonial reticence, and just the type of knowledge/power differentials that this project set out to address (de Saint Laurent, 2018). Along with confusion in language and application, AI-assisted decision-making technologies largely remain the province of powerful economies and as such increase their advantages in trade bargaining. In this paper, we have laid out a set of selection-criteria for identifying AI-enabled technologies applicable to assist and restore some balance to trade negotiations by opening up pathways of information and analysis. The criteria for selecting technologies lead on to broader proposals for information access and analysis that will need to be thoroughly contextualized for each vulnerable economy eventually selected for the real-world project. While the method we have piloted has proved useful for enabling the articulation of known unknown factors influencing the relationship between information access and trading power, and these in turn will better enable sustainable trade negotiations (through power dispersal and sharper market discrimination with more information); another key contribution this multi-disciplinary method offers is the identification of pre-existing unknown knowns (such as alternative sites for data access and/or management, and contextual variables which impact information availability and access, irrespective of AI-assisted technologies).

Acknowledging that AI-assisted information technology access alone will not level the trade bargaining horizon or open up understandings of domestic market sustainability, the scoping and solution exercises suggested some essential pre-conditions: (1) participants in the first group were advantaged as they were familiar with, or had a working knowledge of, current ML technologies; and (2) technical experts in the second had a similar working knowledge of trade bargaining theory, so as to prevent technological solutionism that ignored important social, political, and economic contextual

variables influencing capacities to seek out and understand information asymmetries. Some working technical knowledge connected with contextual sensitivities would ensure that people from both groups are speaking in, not the same language, but unfamiliar ground; and that the articulation of need and sustainability of technological solutions can be as precise as possible. In addition to domestic education and training programs, international organizations have a larger role to play in addressing and reversing the knowledge deficits of technologies in trading and domestic market situations as yet deprived of AI-assisted information pathways. International organizations such as UNCTAD and/or the WTO should thus formulate education policies crafted to enable such productive forms of knowledge exchanges to be initiated before the commensuration of the first scoping exercise. More research can be done here to determine productive forms of such exchanges, and their trajectories. Private sector participants looking for a more resilient global trading and sustainable market future also have a role to play here, as do the large information platform providers in helping to achieve the ESCAP sustainability goals.

The potential for enhancing regional cooperation – in addition to the identification of data pathways – through this method can also serve as a route towards increasing industry standardization and state-to-state data flows, particularly where regional sustainability issues are in the trading conversation. Empowerment approaches beyond nation-state priorities are more likely to achieve scalable deployment and interoperability across countries and can be significantly aided by international coordination bodies such as UNCTAD and/or the WTO working together with standard setting bodies, such as the IEEE³². At this policy level, trading benefit will be viewed as more than only a national concern. Regional approaches to information empowerment and technological capacity building are a realistic recognition that the information which may assist vulnerable economies often knows no jurisdictional boundaries.

32. For example, the IEEE's Data Trading System Initiative. <https://standards.ieee.org/industry-connections/datatradingsystem.html>

References

- de Saint Laurent, C. (2018). In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence. *Europe's Journal of Psychology*, 14(4), 734-737.
- Dikowitz, S. (2014). World Peace Through World Trade. Retrieved from Hinrich Foundation: <https://hinrichfoundation.com/blog/global-trade-world-peace-through-world-trade/>
- Findlay, M. (2017). Law's Regulatory Relevance? Property, Power and Market Economies.
- Fineman, M. A. (2019). Vulnerability and Social Justice. *Valparaiso University Law Review* 53.
- Grcic, J. (2007). Hobbes and Rawls on Political Power. *Ethics & Politics*. Retrieved from <https://core.ac.uk/download/pdf/41174053.pdf>
- Grossman, W. L. (1935). The Legal Philosophy of Roscoe Pound. *Yale Law Review*, 605-618.
- Hardt, M., & Negri, A. (2001). *Empire*. Cambridge: Harvard University Press.
- Humbach, J. A. (2017). Property as Prophecy: Legal Realism and the Indeterminacy of Ownership. *Case Western Reserve Journal of International Law*, 211-225.
- IT Pro team. (2020, March 2). How to measure data quality. Retrieved from ITPro.: <https://www.itpro.co.uk/business-intelligence-bi/29773/how-to-measure-data-quality>
- Reichel, A. (2018). De-growth and Free Trade. Retrieved from <https://www.andrereichel.de/2016/10/18/degrowth-and-free-trade/>
- Stiglitz, J. (2002a). *Globalization and its Discontents*. New York: Penguin, 107.
- Stiglitz, J. (2002b). *Globalization and its Discontents*. New York: Penguin, 246.
- Susaria, A. (2019). The New Digital Divide is People who Opt out of Algorithms and People who. Retrieved from The Telegraph: <https://www.thetelegraph.com/news/article/The-new-digital-divide-is-between-people-who-opt-13773963.php>
- UNCTAD. (2015). General Assembly: Resolution adopted by the General Assembly on 19 December 2014. United Nations. Retrieved from https://unctad.org/en/PublicationsLibrary/ares69d204_en.pdf
- UNTCAD. (2018, October 15). Small economies welcome AI-enabled trade tool but worries remain. Retrieved from UNTCAD: <https://unctad.org/en/pages/newsdetails.aspx?OriginalVersionID=1881>
- UNCTAD. (2019). Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries. United Nation. Retrieved from https://unctad.org/en/PublicationsLibrary/der2019_en.pdf
- UNCTAD. (2019, July 19). Fairer trade can strike a blow against rising inequality. Retrieved from UNCTAD: <https://unctad.org/en/pages/newsdetails.aspx?OriginalVersionID=2154>
- Xiang, A., & Raji, I. D. (2019). On the Legal Compatibility of Fairness Definitions. Retrieved from <https://arxiv.org/abs/1912.00761>

Appendix 1: Hypotheticals

Instructions

Remember your character and your professional location. Reflect on the facts of the following hypotheticals from the perspective of your character and what you understand to be the “knowledge capacity” of trade and sustainability decision-making in your professional location.

Read the following hypotheticals and imagine you are required to participate and to make decisions as instructed with the information provided. At each nominated decision stage, think about what additional information might be useful in making a more effective choice as the factors of the bargain/retention policy are set out.

Clearly, it is difficult to speculate on what you do not know or what is being withheld from you. In this context, common sense as well as experience are useful measures in determining how your decision/bargain would be more empowered through the information available to you. One way of approaching this is to think about the issue/problem that you are confronting, where might be a source of information you currently do not possess, and the form that information might take.

Finally, you are not entirely unfamiliar with information technology. Even though official data, retrieval, and analysis capacity in your professional context is limited, you have a sense of what technological enhancements and information databases those in better resourced administrations and commercial arrangements can access and use to their benefit (and perhaps your detriment). Therefore, you are concerned with information deficit and what information access might enable. You are also interested in how information can be analyzed and applied to make your professional experience more efficient and sustainable.

Hypothetical 1

A large multi-national corporation has commenced discussion with your government to have access to fishing grounds in your territorial waters. Due to the tariff war between several other much larger fishing nations, the price of fish products has grown incrementally in the last economic quarter. The multi-national is also attracted to a trading arrangement with you because your national regulation of fishing practice is neither detailed nor unduly restrictive. In fact, global fishing quotas have largely had little impact on your domestic fishing practice because of its up-until-now subsistence format.

The multi-national has not divulged its intended market for the fish products it would acquire from your waters, but you have some general intelligence that Japan would be a principal third party trader. In Japan, you are aware that the consumer appetite for one particular fish product which is abundant in your waters is high, and prices that can be fetched seem to you to be extraordinary. You have no developed trade arrangement with Japan and you have no detailed understanding of their fish product consumer markets.

The multi-national has also expressed interest in using local labor, the price of which is under-valued due to limited local employment opportunities in the sector. In preliminary meetings, the multi-national has talked of building canning factories for fish processing in two of your major ports where female unemployment is particularly high.

Fish are a dietary staple for many of your citizens living in coastal regions, who practice small scale, indigenous fishing practices. Your fisheries and wildlife department has not done any study on the fish stocks in your territorial waters or on the impact of large-scale commercial fishing on these stocks. You do not have up-to-date information on the multi-national's practices in the harvesting and use of natural resources. In these negotiations, you would be dealing with a subsidiary of the larger multi-national set up specifically for this trading exercise and registered in the Republic of Ireland for beneficial taxation concessions.

You have been asked:

- a) To further the preliminary negotiations with the multi-national;
- b) To oversee an environmental impact assessment of the proposal;
- c) To draft conditions under which specific trade negotiations might be structured;
- d) To address concerns from local indigenous fishing communities.

Hypothetical 2

A consortium of foreign investors has approached your government with the intention of structuring and implementing some foreign direct investment (FDI) infra-structure projects in your country. The consortium consists of a major Chinese banking group, an international construction company, a major power generator, and a telecommunications provider. The types of projects being discussed are very attractive to your under-capitalized transport and communications sector.

A condition of the foreign direct investment portfolio is that your government signs up to various loan agreements offered by the Chinese bank. As a condition of the loans, your government will agree to having any disputes arising between your state and the consortium arbitrated in China under Chinese commercial law.

The international construction company will design and build a new dam over a large natural river system. Water resources are a major concern for your country. Because of what they refer to as 'technology considerations', the construction company intends only to use its own imported labor.

Your state is in desperate need of power generating facilities. The major power generator in the consortium is happy to finance the construction and operation of a nuclear power plant within your territory, provided that you allow half of the power generated in that grid to be independently traded by the consortium into neighboring states. In addition, the consortium wants your government to cease discussions with another neighbor states for the shared construction of wind farms on your border.

The telecommunications provider will invest in 5G technologies throughout your state. Most of your communication capacity at present is not fully 4G compliant. There have been concerns expressed in your business community that such a rapid convergence into 5G might produce significant secondary costs through unnecessary technological obsolescence. Furthermore, talk from the telecommunications about linking your 5G capacity to developments in the Internet of Things (IoT) in China, seem obscure and unclear.

You have been asked:

- a) To further the preliminary negotiations with the consortium;
- b) To oversee an economy-wide evaluation of the impact of the proposed FDI;
- c) To draft conditions under which specific investment negotiations might be structured;
- d) To address concerns from local businesses such as the domestic power provider, domestic telcos, and local trade unions regarding medium-term sustainability issues.

Hypothetical 3

In an effort to improve your trade imbalance, your government over recent decades has implemented an agricultural policy of transition from subsistence to cash cropping. In particular, palm oil plantations have been incentivized and major regional companies have invested in concessions for palm oil production. A political consequence has been push-back from smaller farmers who are unable to match the economies of scale of the bigger plantations. To confront this resistance, the government has operated a subsidy system to encourage small farmers to cash crop, and to compensate for their market disadvantage.

Both the bigger producers and the small farmers employ slash-and-burn clearing techniques, which has caused air pollution with associated damage to the health of the domestic population and neighboring states.

The government is worried about its growing dependence on a single export crop, when global market vulnerability is difficult to predict. Entrepreneurs from Canada, which recently legalized the growing and use of marijuana, are in discussions with your government to invest in major hemp farms in your country for export back to Canada and California, where they say the market is expanding. Governments in your region with tough anti-drug laws have lobbied your government against the initiative. Marijuana is currently a prescribed drug in your jurisdiction, but popular opinion would be tolerant of decriminalization for medical and economic reasons.

The Canadian investors have also indicated – to improve the attractiveness of their agricultural intentions – to bring with them a significant research and development investment that could stimulate the growth of a generic drug industry in your country; namely, processing the medical constituents of marijuana. This industry would, they say, offers employment mobility for semi-skilled workers currently occupied in low-paid sweat shop garment-making, which is another diminishing domestic export industry here.

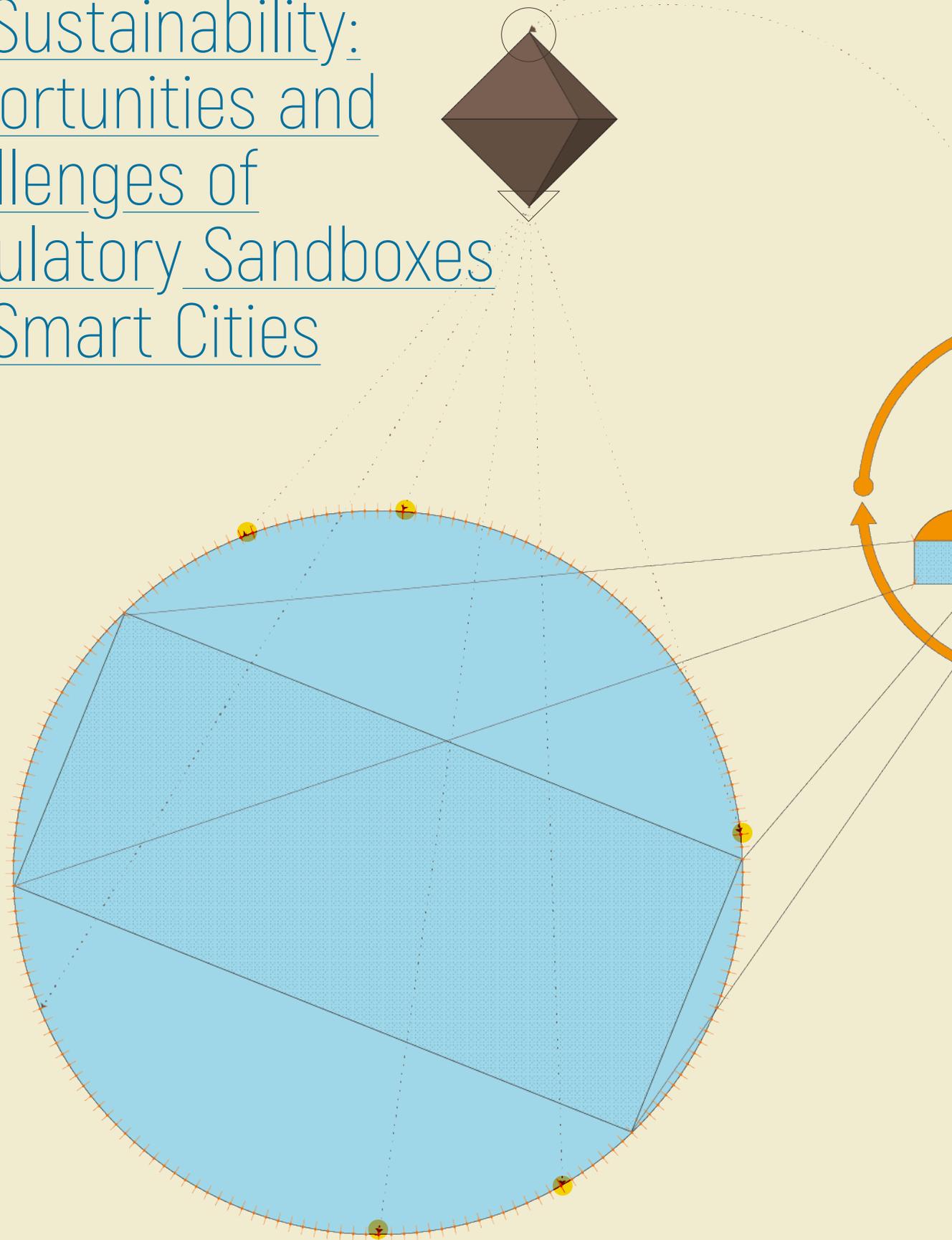
You have been asked:

- a) To further the preliminary negotiations with the Canadian investors;
- b) To oversee a comparative environmental impact assessment of the proposal relative to existing cash cropping practices;
- c) To draft conditions under which investment negotiations might be structured;
- d) To address concerns on the relationship between trade and regional foreign policy.

Governing Data-driven Innovation for Sustainability: Opportunities and Challenges of Regulatory Sandboxes for Smart Cities

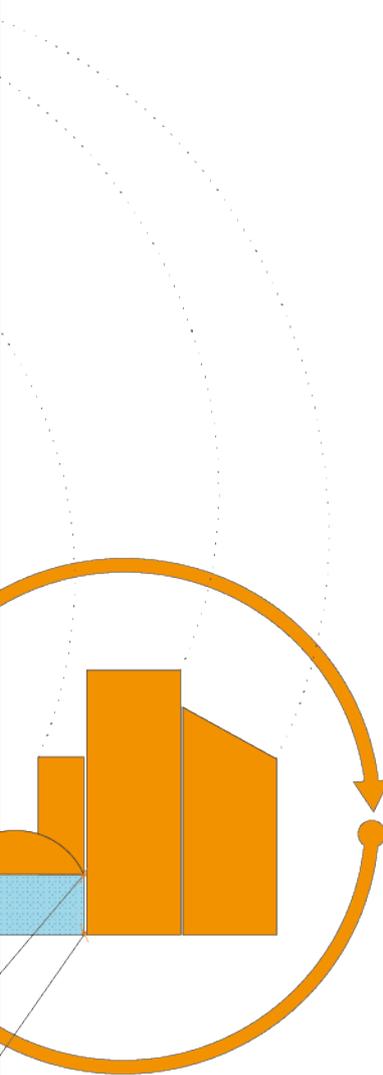
Masaru Yarime¹

Division of Public Policy,
The Hong Kong University of
Science and Technology



1. I would like to thank Gleb Papishev for his assistance in preparing this report.

Abstract



Data-driven innovation plays a crucial role in tackling sustainability issues. Governing data-driven innovation is a critical challenge in the context of accelerating technological progress and deepening interconnection and interdependence. AI-based innovation becomes robust by involving the stakeholders who will interact with the technology early in development, obtaining a deep understanding of their needs, expectations, values, and preferences, and testing ideas and prototypes with them throughout the entire process. The approach of regulatory sandboxes will particularly play an essential role in governing data-driven innovation in smart cities, which inevitably faces a difficult challenge of collecting, sharing, and using various kinds of data for innovation while addressing societal concerns about privacy and security. How regulatory sandboxes are designed and implemented can be locally adjusted, based on the specificities of the economic and social conditions and contexts, to maximize the effect of learning through trial and error. Regulatory sandboxes need to be both flexible to accommodate the uncertainties of innovation, and precise enough to impose society's preferences on emerging innovation, functioning as a nexus of top-down strategic planning and bottom-up entrepreneurial initiatives. Data governance is critical to maximizing the potential of data-driven innovation while minimizing risks to individuals and communities. With data trusts, the organizations that collect and hold data permit an independent institution to make decisions about who has access to data under what conditions, how that data is used and shared and for what purposes, and who can benefit from it. Alternatively, a data linkage platform can facilitate close coordination between the various services provided and the data stored in a distributed manner, without maintaining an extensive central database. The data governance systems of smart cities should be open, transparent, and inclusive. As the provision of personal data would require the consent of people, it needs to be clear and transparent to relevant stakeholders how decisions can be made in procedures concerning the use of personal data for public purposes. The process of building a consensus among residents needs to be well-integrated into the planning of smart cities, with the methodologies and procedures for consensus-building specified and institutionalized in an open and inclusive manner. It is also essential to respect the rights of those residents who do not want to participate in the data governance scheme of smart cities. As APIs play a crucial role in facilitating interoperability and data flow in smart cities, open APIs will facilitate the efficient connection of various kinds of data and sophisticated services. International cooperation will be critically important to develop common policy frameworks and guidelines for facilitating open data flow while maintaining public trust among smart cities across the globe.

Introduction

Data-driven innovation plays a crucial role in tackling sustainability challenges such as reducing air pollution, increasing energy efficiency, eliminating traffic congestion, improving public health, and maintaining resilience to accidents and natural disasters (Yarime, 2017). These multifaceted challenges, which are interconnected and interdependent in complex ways, require the effective use of various kinds of data concerning environmental, economic, social, and technological aspects that are increasingly available through sophisticated equipment and devices in smart cities. Innovation based on artificial intelligence (AI) can make the best use of these data to accelerate learning and improve performance. It is of critical importance to establish adaptive governance systems that allow experimentation and flexibility to deal with the uncertainty and unpredictability of technological change, while addressing societal concerns such as security and privacy incorporating local contexts and conditions. Novel forms of technology governance, such as testbeds, living laboratories, and regulatory sandboxes, are required for policymakers to address the evolving nature of data-based innovation.

In this paper, we examine key opportunities and challenges in the governance of data-driven innovation in the context of smart cities. First, we discuss the major characteristics of data-driven innovation and highlight the importance of learning and adaptation through the actual use of technologies in real situations. Next, we examine the approach of regulatory sandboxes to facilitate innovation by taking previous examples of introducing them to the field of finance and other sectors with their experiences and implications. Then we consider emerging cases of applying regulatory sandboxes to stimulate novel technologies utilizing AI in cyber-physical systems such as drones, autonomous vehicles, and smart cities. Finally, we discuss critical challenges in designing and implementing regulatory sandboxes for AI-based innovation, with a particular focus on

data governance. Implications are explored for data governance to promote the collection, sharing, and use of data for innovation while taking appropriate measures to address societal concerns, including safety, security, and privacy. Recommendations for policymakers are considered to facilitate the engagement of relevant stakeholders in society so that various kinds of data collected in smart cities are appropriately used to govern innovation based on AI.

Characteristics of Data-driven Innovation

The emergence of data-driven innovation based on the rapid advancement in the Internet of Things (IoT) and AI creates exciting opportunities as well as considerable challenges in promoting societal benefits while regulating the associated risks. As a vast amount of diverse kinds of data is increasingly available from various sources that were not previously accessible, a wide range of sectors are currently undergoing significant transformation. In energy, smart grid systems lower costs, integrate renewable energies, and balance loads. In transportation, dynamic congestion-charging systems adjust traffic flows and offer incentives to use park-and-ride schemes, depending upon real-time traffic levels and air quality. Car-to-car communication can manage traffic to minimize transit times and emissions, and eliminate road deaths from collisions (Curley, 2016). The speed of technological advancement is accelerating, and those technologies that used to be separate are increasingly interconnected and interdependent with one another, creating a significant degree of uncertainty in their impacts and consequences.

The process of data-driven innovation has three key components: data collection, data analysis, and decision making (Organisation for Economic Co-operation and Development, 2015a). Data-driven innovation critically depends on the efficient and effective collection, exchange, and sharing of large

amounts of high-quality data. New technologies such as drones, IoT, and satellite images can now provide vast amounts of data that were not previously available or accessible before. The big data collected through various sources and challenges are analyzed by applying data science. Sophisticated methodologies and tools are increasingly possible due to the recent technological advancement in AI, particularly the rapid progress in machine learning. For decision making, it is critical to integrate the findings of data analytics with the domain expertise that would be specific to the sector in which you are involved, such as energy, health, or transportation. Increasingly, cyber systems are merging with physical machines and instruments as in manufacturing, and such cyber-physical systems are particularly important in dealing with sustainability issues in the context of smart cities.

Data-driven innovation is accelerated by deriving new and significant insights from the vast amount of data generated during the delivery of services every day. Hence training, the ability to learn from real-world use and experience, and adaptation, the capability to improve the performance, would be key to creating data-driven innovation (Food and Drug Administration, 2019). The development of cyber-physical systems such as smart cities is facilitated through the ready availability of and accessibility to data, as well as its mutual exchange and sharing with stakeholders in different sectors. Unlike the traditional model of innovation, which tends to rely on closed, well-established relationships between enterprises in a specific industry, the new mode of data-driven innovation requires open, dynamic interactions with stakeholders possessing and generating various kinds of data. Close cooperation and collaboration on data become crucial in the innovation process, from the development of novel technologies to deployment through field experimentation and legitimation in society.

There are difficult challenges to policymakers in facilitating data-driven innovation in cyber-physical systems. The speed of technological change of

AI is remarkably fast, which has been particularly demonstrated in the case of image recognition (Russakovsky, Deng, Su, Krause, Satheesh, Ma, Huang, Karpathy, Khosla, Bernstein, Berg & Fei-Fei, 2015). That leads to remarkable progress in the performance of AI and, at the same time, accompanies a significant degree of uncertainty in consequences and side effects. Various kinds of technologies are increasingly interconnected and interdependent through data exchange and sharing among multiple sectors, such as energy, buildings, transportation, and health. These characteristics make it difficult to explain or understand the process of innovation and contribute to giving rise to a widening gap between technological and institutional changes. It is critical to establish a proper system to govern data-driven innovation in the context of accelerating technological progress and deepening interconnection and interdependence. New policy approaches are required to stimulate data-driven innovation in cyber-physical systems by facilitating coordination and integration of emerging technologies while addressing societal concerns such as safety, security, and privacy.

As the introduction of AI systems is relatively new, our understanding of the behavior of such systems in real-life situations is still minimal. As machines powered by AI increasingly mediate our economic and social interactions, understanding the behavior of AI systems is essential to our ability to control their actions, reap their benefits, and minimize their harms (Rahwan, Cebrian, Obradovich, Bongard, Bonnefon, Breazeal, Crandall, Christakis, Couzin, Jackson, Jennings, Kamar, Kloumann, Larochelle, Lazer, McElreath, Mislove, Parkes, Pentland, Roberts, Shariff, Tenenbaum & Wellman, 2019). AI systems cannot be entirely separate from the underlying data on which they are trained or developed. Hence it is critical to understand how machine behaviors vary with altered environmental inputs, just as biological agents' behaviors vary depending on the environments in which they exist. Our understanding of the behavior of AI-based systems can benefit from an experimental examination of human-machine interactions in real-world settings.

The experience of using an AI system in clinics in Thailand for the detection of diabetic eye disease is one of the few cases that provide valuable lessons and implications (Beede, Baylor, Hersch, Iurchenko, Wilcox, Ruamviboonsuk & Vardoulakis, 2020). While deep learning algorithms promise to improve clinician workflows and patient outcomes, these gains have not been sufficiently demonstrated in real-world clinical settings. The Ministry of Health in Thailand has set a goal to screen 60% of its diabetic population for diabetic retinopathy (DR), which is caused by chronically high blood sugar that damages blood vessels in the retina. Reaching this goal, however, is a challenge due to a shortage of clinical specialists. That limits the ability to screen patients and also creates a treatment backlog for those found to have DR. Thus, nurses conduct DR screenings when patients come in for diabetes check-ups by taking photos of the retina and sending them to an ophthalmologist for review. A deep learning algorithm has been developed to provide an assessment of diabetic retinopathy, avoiding the need to wait weeks for an ophthalmologist to review the retinal images. This algorithm has been shown to have specialist-level accuracy for the detection of referable cases of diabetic retinopathy. Currently, there are no requirements for AI systems to be evaluated through observational clinical studies, nor is it common practice. That is problematic because the success of a deep learning model does not rest solely on its accuracy, but also on its ability to improve patient care.

This experience provides critical recommendations for continued product development and guidance

on deploying AI in real-world scenarios (Beede, 2020). The functioning of AI systems in healthcare is affected by workflows, system transparency, and trust, as well as environmental factors such as lighting which vary among clinics and can impact the quality of images. AI systems need to be trained to handle these situations. An AI system might conservatively determine some images having blurs or dark areas to be ungradable because they might obscure critical anatomical features required to provide a definitive result. On the other hand, the gradability of an image may vary depending on a clinician's experience or physical set-up. Any disagreements between the AI system and the clinician can create problems. The research protocol has been subsequently revised, and now eye specialists review such ungradable images alongside the patient's medical records, instead of automatically referring patients with ungradable images to an ophthalmologist. This helped to ensure a referral was necessary and reduced unnecessary travel, missed work, and anxiety about receiving a possible false-positive result. In addition to evaluating the performance, reliability, and clinical safety of an AI system, we also need to consider the human impacts of integrating an AI system into patient care. The AI system could empower nurses to confidently and immediately identify a positive screening, resulting in quicker referrals to an ophthalmologist.

This case highlights that, in addition to the accuracy of the algorithm itself, the interactions between end-users and their environment determine how a new system based on AI will be implemented, which cannot always be controlled through careful planning. Even

when a deep learning system performs a relatively straightforward task, for example, just analyzing retinal images, organizational or socio-environmental factors are likely to impact the performance of the system. Many environmental factors that negatively impact model performance in the real world might be reduced or eliminated by technical measures, such as through lighting adjustments and camera repairs. However, these types of modifications could be costly and even infeasible in low-resource settings, making it even more critical to engage with contextual phenomena from the start. AI-based innovation becomes robust by involving the stakeholders who will interact with the technology early in development, obtaining a deep understanding of their needs, expectations, values, and preferences, and testing ideas and prototypes with them throughout the entire process.

The findings of the actual case of implementing AI-based innovation provide useful implications for technology policy and governance. As policy makers are required to respond to technological change in real-life situations, technology governance becomes an integral part of the innovation process itself to steer emerging technologies towards better collective outcomes. Governments need to anticipate significant changes induced by autonomous vehicles, drone technologies, and widespread IoT solutions, as well as to consider their implications for public policy. AI technologies offer opportunities to improve economic efficiency and quality of life, but they also bring many uncertainties, unintended consequences, and risks. As such, this calls for more anticipatory and participatory modes of governance (OECD, 2018).

Anticipatory governance acts on a variety of inputs to manage emerging knowledge-based technologies and the missions built upon them, while such management is still possible (Guston, 2014). It requires government foresight, engagement, and reflexivity to facilitate public acceptance of new technologies, while at the same time assessing, discussing, and preparing for their intended and unintended economic and societal effects. Anticipatory approaches can help explore, consult widely on, and steer the consequences of innovation at an early stage and incorporate public values and concerns, mitigating potential backlash against technology. Traditional policy tools would not be able to deal with situations where the future direction of technological innovation cannot be determined. In contrast, new policy tools such as regulatory sandboxes emphasize the benefits of environments that facilitate learning to help understand the regulatory implications and responses to emerging technologies. Participatory approaches can provide a wide range of stakeholders, including citizens, with adequate opportunities to appraise and shape technology pathways (OECD, 2018). These practices can help ensure that the goals, values, and concerns of society are continuously enforced in emerging technologies, and shape technological designs and trajectories without unduly constraining innovators. This will contribute to supporting efforts to promote responsible innovation, which has integrated dimensions of anticipation, reflexivity, inclusion, and responsiveness (Stilgoe, Owen & Macnaghten, 2013).

The Approach of Regulatory Sandboxes

The approach of regulatory sandboxes has recently been proposed to stimulate innovation by allowing experimental trials of novel technologies and systems that cannot currently operate under the existing regulations by specifically designating geographical areas or sectoral domains. Regulatory sandboxes provide a limited form of regulatory waiver or flexibility for firms to test new products or business models with reduced regulatory requirements, while preserving some safeguards to ensure appropriate consumer protection (Organisation for Economic Co-operation and Development, 2019). Potential benefits include facilitating greater data availability, accessibility, and usability for innovators, and reducing the time and cost of getting innovative ideas to market by reducing regulatory constraints and ambiguities (Financial Conduct Authority, 2015). The approach aims to provide a symbiotic environment for innovators to test new technologies and for regulators to understand their implications for industrial innovation and consumer protection. The aim is to help identify and better respond to regulatory breaches by enhancing flexibility and adjustment in regulations, which would be particularly relevant in highly regulated industries, such as the finance, energy, transport, and health sectors.

Regulatory sandboxes have initially been introduced to the financial sector in efforts to encourage fintech by providing a regulatory safe space for innovative financial institutions and activities underpinned by technology (Zetsche, Buckley, Barberis & Arner, 2017). While the sandbox creates an environment for businesses to test products with less risk of being punished by the regulator for non-compliance, regulators require applicants to incorporate appropriate safeguards to insulate the market from risks of their innovative business. In early 2016, the Financial Conduct Authority (FCA) of the UK initiated a fintech regulatory sandbox to encourage innovation in the field of financial technology. The sandbox aimed to

provide the conditions for businesses to test innovative products and services in a controlled environment without incurring the regulatory consequences of pilot projects (Financial Conduct Authority, 2015). A fintech supervisory sandbox was also launched by the Hong Kong Monetary Authority in September 2016, followed by other fintech sandboxes in Australia, Canada, and Singapore. The concept has also been embraced by a growing number of developing world regulators as well.

There are some lessons learned from the experience of regulatory sandboxes in fintech (Financial Conduct Authority, 2017). Working closely with the FCA has allowed firms to develop their business models with consumers in mind and mitigate risks by implementing appropriate safeguards to prevent harm. A set of standard safeguards have been put in place for all sandbox tests. All firms in the sandbox are required to develop an exit plan to ensure that the test can be terminated whenever it is necessary to stop the potential harm to participating consumers. The sandbox has allowed the agency to work with innovators to build appropriate consumer protection safeguards into new products and services.

The approach of regulatory sandboxes has gone beyond the field of finance and has been applied in other sectors involving cyber-physical systems, which more directly concern safety, human health, and public security. In the energy sector, the Office of Gas and Energy Markets (Ofgem) of the UK started their Innovation Link service in February 2017 as a one-stop shop offering rapid advice on energy regulation to businesses looking to launch new products or business models (Office of Gas and Electricity Markets, 2018a). When regulatory barriers prevent launching a product or service that would benefit consumers, a regulatory sandbox can be granted to enable a trial.

The Energy Market Authority (EMA) in Singapore also launched a regulatory sandbox in October 2017 to encourage experimentation of new products and services in the electricity and gas sectors (Energy Market Authority, 2017). EMA, as the industry regulator, assesses the impact of new products and services before deciding on the appropriate regulatory treatment. Innovators submit their ideas to EMA for testing, and a successful application allows the plan to be applied in the market while being subject to relaxed regulatory requirements. Safeguards such as limiting the duration of the trial or the maximum number of consumers can be introduced to minimize risks to consumers and industry. The evaluation criteria when applying for the regulatory sandbox include using technologies or products in an innovative way, addressing a problem or bringing benefits to consumers or the energy sector, requiring some changes to existing rules, and having assessed and mitigated foreseeable risks. The regulatory sandbox complements ongoing energy research and development (R&D) initiatives by providing a platform for R&D projects to be tested on a broader scale in the country.

The experience of introducing regulatory sandboxes to the energy sector offers a number of lessons and implications. Ofgem's officials spent time talking to innovators to understand their business and to locate and interpret the rules that affected them. Through an iterative process, they effectively worked with innovators to co-create feasible sandbox trials (Office of Gas and Electricity Markets, 2018b). It was not always clear to innovators what they could or could not do, nor always easy for them to find rules or interpret them. Hence advice from the agency helped the innovators figure out which regulations would be relevant for their technologies or services. Sometimes proposals were not allowed for trials, as some institutional requirements, including industry norms, systems, charging arrangements, codes, and licenses, became obstacles. While the sandbox was introduced to facilitate time-limited trials with the

temporary relaxation of rules, most innovators would like to continue to operate after the test and to see the experience of regulatory sandboxes used to change the existing policies and regulations.

The approach of regulatory sandboxes can play an essential role in governing data-driven innovation, which inevitably faces a difficult challenge of collecting, sharing, and using various kinds of data for innovation while addressing societal concerns about privacy and security. The Information Commissioner's Office (ICO) in the UK has recently introduced a regulatory sandbox that is designed to support start-ups, SMEs, and large organizations across private, public, and voluntary sectors. The condition is that they use personal data to develop products and services which are innovative and have demonstrable public benefits (Information Commissioner's Office, 2019). The regulatory sandbox enables participants to consider how they use personal data in their projects, as well as provides some comfort from enforcement action and increases public reassurance that innovative products and services are not in breach of data protection legislation. As these products and services are considered to be on the cutting edge of innovation and operating in particularly challenging areas of data protection, there is a significant extent of uncertainty about adequately complying with the relevant regulations. Participants in the regulatory sandbox can become use cases, and, subsequently, the ICO would be able to revise public guidance and provide necessary resources for compliance.

An important issue in designing and implementing regulatory sandboxes is how to manage regulatory arbitrage. Regulatory sandboxes aim to stimulate innovation by relaxing relevant regulations so that entrepreneurs can experiment with novel technologies without being constrained too much by the existing regulatory environment. This creates opportunities for regulatory arbitrage, which refers collectively to the strategies that can be used to achieve an economically equivalent outcome to a regulated

activity while avoiding the legal constraints (Fleischer, 2010). It is a legal planning technique used to avoid regulatory requirements such as taxes, accounting rules, securities disclosure, with other requirements such as safety and privacy also possibly being included. Jurisdictionally speaking, regulatory arbitrage means that a firm chooses a location where a more favorable regulatory treatment is available to its business activities (Allen, 2019). While national borders do not constrain the development and deployment of AI-based products and services, regulatory sandboxes have only been created at national or sub-national levels. This discrepancy can lead to what is known as the race to the bottom, a phenomenon where jurisdictions compete to lower their regulatory standards in order to attract innovative companies, which could potentially result in negative consequences on consumer protection with regard to safety and privacy.

The challenges of regulatory arbitrage and the race to the bottom can be tackled if the regulators in different locations can coordinate with one other to share the information necessary to formulate appropriate policy measures and commit to agreements to apply consistently high regulatory standards (Allen, 2020). Regulators, however, have their specific policy preferences and strong incentives to keep information within individual regulatory sandboxes, rather than share it with other sandboxes in different locations. Social license and the bundling of laws and resources could work as constraining forces on regulatory arbitrage (Pollman, 2019). Aggressive regulatory arbitrage can erode social license and create a costly environment for sustainable operation, especially when social costs are widely recognized in the community. Also, as an opportunity for regulatory arbitrage would arise not in isolation but within a system of laws, and in light of other considerations such as investment capital and workforce talent, the bundling of relevant laws and regulations would leave less room available for regulatory arbitrage. If the existing laws create a regulatory environment that is prohibitive to a particular type of innovation, companies may try to

focus on changing the legal environment rather than merely arbitrage regulatory differences. A complex set of factors and considerations would influence decisions about regulatory arbitrage, which includes transparency of information to the public and the ability of a company to mobilize its resources for regulatory change.

Moving in a more positive direction, an increasing number of enterprises actually try to advance innovative technologies by strategically taking regulatory arbitrage. One example is Cyberdyne, a Japanese company that developed a medical and healthcare robot, HAL (Ikeda and Iizuka, 2019). Under the Japanese product classification system, HAL could be categorized as a medical device or an assistive device, each of which would be regulated by different institutions. Although the company initially planned to commercialize the robot as a medical device with public medical insurance coverage, that required the product to comply with rigorous medical safety regulations with clinical trials. Considering the regulatory environment, Cyberdyne first chose to commercialize HAL as an assistive device, which usually requires proof of safety, certified by a third party on voluntary terms. The Robot Safety Centre, a public institution located in the Tsukuba International Strategic Zone, Tokku, supported the company to conduct the necessary testing and produce evidence for proof of safety. During this process the company was able to accumulate experiences to improve the product, which was eventually certified by the Japan Quality Assurance Organization and commercialized as an assistive device.

On the other hand, Cyberdyne chose to commercialize HAL as a medical device in Germany first (Iizuka and Ikeda, 2019). From the beginning there was an expectation that it would take a long time to receive an approval from the Ministry of Health, Labour and Welfare (MHLW) in Japan because there was no precedent product similar to the new robot. In Germany, in contrast, a new health device like HAL

is categorized solely as a medical device strictly by its function regardless of its risk levels on safety. As the review of medical devices is certified by a private certification body, the procedure is codified, open, and transparent, and the time required for approval of new medical devices is substantially less than in Japan. HAL has been certified as a medical device in Germany and subsequently commercialized in Europe. After that, the robot was approved by the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan and commercialized with public insurance coverage.

At the same time, Cyberdyne also engaged in developing ISO standards for the safety of personal care robots including healthcare robots (Iizuka & Ikeda, 2019). As there had not been robots like HAL before, there was no regulation in place to protect users, and international standards were considered to be crucial for establishing confidence in these products. Also, while these new standards can guarantee the company an early-mover advantage with the global recognition of its brand, they level the playing field for new entrants to the emerging industry. As Cyberdyne was already developing personal care robotics and was experimenting with prototype safety measures, a set of evidence created during this process became a basis to establish ISO standards on robotics safety.

This case demonstrates a possibility that regulatory arbitrage can actually function to promote innovation. As a start-up with limited resources, Cyberdyne did not attempt to directly influence the relevant regulations. The company instead tried to cope with the regulatory obstacle by commercializing the new robot in the domestic market as an assistive device first and further developing the technology as a medical device overseas. The company also participated in setting up the institutional environment in which the new product is recognized properly. Hence regulatory arbitrage can also mean that enterprises strategically take advantage of differences in regulatory systems to develop and commercialize innovative products while contributing to establishing institutions to facilitate market creation.

Cases of Regulatory Sandboxes for AI-based Innovation

For smart city development, demonstration projects play an increasingly crucial role in testing novel technologies and raising awareness among the general public. These projects are mainly aimed at examining promising but unproven technologies concerning various aspects of cities, including energy, transportation, buildings, health, environment, and infrastructure. Existing policies and regulations, however, may not necessarily be able to properly deal with certain unexpected novel features of technologies. Hence entrepreneurs and innovators would have difficulties in conducting field testing of emerging technologies on the ground, particularly when other stakeholders, including local communities and residents, are involved. Regulatory sandboxes can relax or adjust some of the relevant regulations so that these new technologies can be tested for actual adoption and use. How regulatory sandboxes are designed and implemented can be locally adjusted, based on the specificities of the economic and social conditions and contexts, to maximize the effect of learning through trial and error. Various types of new promising technologies can be verified, adopted, and integrated, effectively improving technological performance, reliability, and integration, as well as contributing to cost reduction.

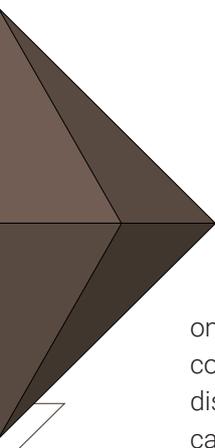
In particular, regulatory sandboxes can improve the understanding of how AI systems may react in specific contexts and satisfy human needs. As AI-based innovation involves rapid technological change, uncertain market development, and diverse social norms, there are many economic, ethical, and legal issues comprised of various interests and preferences. It is necessary to have a regulatory framework that is flexible enough to accommodate the uncertainties of innovation and, at the same time, clear enough to impose society's preferences on emerging innovation. This requires a specific form of governance that incorporates both elements of top-

down legal framing and bottom-up empowerment of individual actors (Pagallo, Aurucci, Casanovas, Chatila, Chazerand, Dignum, Luetge, Madelin, Schafer & Valcke, 2019). Regulatory sandboxes can function as a nexus of top-down strategic planning and bottom-up entrepreneurial initiatives.

The current regulations in the fields of autonomous vehicles, drones, and medical devices show that rules on AI are significantly dependent upon the context of locations and sectors (Pagallo, Aurucci, Casanovas, Chatila, Chazerand, Dignum, Luetge, Madelin, Schafer & Valcke, 2019). In the case of the EU, for example, in addition to the rules on data protection, the testing and use of self-driving cars needs to comply with a complex legal network involving three directives and one regulation: Council Directive 85/374/EEC on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products; Directive 1999/44/EC on certain aspects of the sale of consumer goods and associated guarantee, such as repair and replacement, and price reduction and termination; Directive 2009/103/EC relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability; and Regulation 2018/858 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components, and separate technical units intended for such vehicles. The testing and use of drones requires compliance with one regulation, Regulation (EU) 2018/1139 on common rules in the field of civil aviation and establishing a European Aviation Safety Agency, and two European Commission implementing and delegated acts, Delegated Regulation 2019/945 and the Implementing Regulation 2019/947, in addition to several opinions and guidelines of the European Aviation Safety Agency (EASA). Medical devices based on AI need to deal with contractual and tort liability in national regulations of the EU member states.

Given the rapid progress and unpredictable evolution of AI-based innovation, some countries have established special deregulated zones as living labs to allow testing and experimentation of new technologies in actual fields. In Japan, the National Strategic Special Zones system was introduced in 2013 to enhance economic growth by implementing regulatory reforms. So far, ten areas have been designated as special zones, and more than 60 reforms have been realized, with over 350 projects currently ongoing as a result of these regulatory reforms (Secretariat for the Promotion of Regional Development, 2019). In these special zones, regulatory exceptions have been introduced without amending the laws by taking into account specific local circumstances, and municipalities and private companies have proposed voluntary plans. Specifically targeting self-driving vehicles, in October 2017, the government introduced the National Strategic Special Zones for Level 4 Automated Vehicles Deployment Project on public roads. With the aim of establishing social and legal systems for future technological development, public road safety demonstration experiments were conducted. Based on the experience of building these special zones, the Japanese government initiated a new framework for regulatory sandboxes in March 2018, covering financial services, healthcare industry, mobility, and transportation.

In Singapore, the Road Traffic Act was amended in February 2017 to recognize that a motor vehicle need not have a human driver. The Minister for Transport is able to create new rules on trials of autonomous vehicles, acquire the data from the trials, and set standards for autonomous vehicle designs (Taeiagh & Lim, 2019). A five-year regulatory sandbox was created to ensure that innovation is not stifled, and the government intends to enact further legislation in the future. Autonomous vehicles must pass safety assessments, robust plans for accident mitigation must be developed before road testing, and the default requirement for a human driver can be waived



once the autonomous vehicle demonstrates sufficient competency to the Land Transport Authority. After displaying higher competencies, autonomous vehicles can undergo trials on increasingly complex roads.

In 2017, the United States Federal Aviation Administration (FAA) launched the Unmanned Aircraft System (UAS) Integration Pilot Program (IPP), with fixed-term regulatory exemptions and adaptive regulations, to test the safe application of drones (Federal Aviation Administration, 2019). The program has helped the Department of Transportation and FAA develop new rules that support more complex low-altitude operations by addressing security and privacy risks and accelerating the approval of operations that currently require special authorizations. Ten public-private partnerships have been chosen to test the use of unmanned aerial vehicles (UAV), drones, in potentially useful ways that are currently illegal under federal law without a waiver (Boyd, 2018). The program encouraged applicants to submit proposals for test cases that would obtain data that could be applied to broader use cases, with the understanding that the Department of Transportation and FAA would waive certain restrictions to make these projects viable. The IPP Lead Participants are evaluating a host of operational concepts, including night operations, flights over people and beyond the pilot's line of sight, package delivery, detect-and-avoid technologies, and the reliability and security of data links between pilot and aircraft, with potential opportunities for application in commerce, photography, emergency management, agricultural support, and infrastructure inspections.

In Germany, the energy sector is emphasized to encourage innovative solutions for a future energy system based on renewable energy and higher energy efficiency through digitalization. The Economic Affairs Ministry has set up a large-scale regulatory sandbox entitled Smart Energy Showcases – Digital Agenda for the Energy Transition (SINTEG). It offers temporary

spaces in which solutions for technical, economic, and regulatory challenges relating to energy transition can be developed and demonstrated (Federal Ministry for Economic Affairs and Energy, 2019). Moreover, a scheme for regulatory sandboxes has been established to test technical and non-technical innovations in real life and on an industrial scale in critical areas of energy transition. As the smart cities project aims to test various possibilities for digitalization and ensure a good fit with sustainable and integrated urban development, the Federal Ministry of the Interior, Building, and Community has been funding the project since 2019.

For autonomous vehicles, the Federal Ministry of Transport and Digital Infrastructure (BMVI) established the Digital Motorway Test Bed to allow testing of the latest automated driving technology in a real-life setting. The Hamburg Electric Autonomous Transportation project (HEAT) investigates how fully autonomous or self-driving electric minibuses can be safely deployed to transport passengers on urban roads. Since the test vehicles are powered vehicles with highly or fully automated driving functions, the implementation of the project and registration of the cars necessitates applications according to the German Road Vehicles Registration and Licensing Regulations, with exemptions. Regulatory sandboxes can also be designed as testbeds for broad-based participation. The Baden-Württemberg Autonomous Driving Testbed is a regulatory sandbox for mobility concepts that permits companies and research establishments to test technologies and services in the field of connected and automated driving. The combination of various elements of relevance to mobility and the consortium of scientific and municipal partners creates a platform on which key insights and momentum can be gained for the ongoing development of legislation and policy for autonomous driving.

The approach of regulatory sandboxes has been identified as an essential policy instrument for promoting responsible innovation in the national strategy for AI of Norway (Norwegian Ministry of Local Government and Modernisation, 2020). In this strategy, the concept refers to legislative amendments that allow trials within a limited geographical area or period, as well as more comprehensive measures in areas where the relevant supervisory authority needs close monitoring and supervision. The government has established regulatory sandboxes in the field of transportation in the form of legislative amendments that allow testing activities. An act to enable pilot projects on autonomous vehicles came into force in January 2018. Maritime authorities established the first test bed for autonomous vessels in 2016, and two more test beds have been approved since then. In 2019 parliament adopted a new Harbours and Fairways Act, which permits autonomous coastal shipping. Such permission allows sailing in specific fairways, subject to compulsory pilotage or in areas where no pilotage services are provided. Where pilot projects deviate from applicable laws and regulations, they can be conducted with statutory authority in special rules. Alternatively, under the Pilot Schemes in Public Administration Act, public administration can apply to the Ministry of Local Government and Modernisation to deviate from laws and regulations to test new ways of organizing their activities or performing their tasks for a period of up to four years.

In the UK, technology suppliers and their National Health Service (NHS) partners who were delivering machine learning applications in diagnostic pathways have begun work on a regulatory sandbox (Care Quality Commission, 2020). The Care Quality Commission (CQC) formed a team with members from across different functions, as well as a governance committee to oversee the work. The National Institute for Clinical Excellence (NICE), the Medicines and Healthcare Products Regulatory Agency (MHRA), and the NHSX – a joint unit between the NHS and the Department of Health and Social Care to drive the digital transformation of health care – were also included as government partners in this sandbox. They have been working to explore new guidance for NHS providers on AI systems with the

Information Commissioner's Office. The first output from the regulatory sandbox process is a common understanding of what should be present to help deliver high-quality care when using machine learning applications in clinical diagnostics. Developing this shared view of quality with people who use services, providers, technology suppliers, and system partners has been the basis of their work in the sandbox.

In Europe, deregulated special zones have mainly been applied in the fields of self-driving cars and drones. The Swedish government sponsored the world's first large-scale autonomous driving pilot project in 2016. In Belgium, the first special zone for the testing of drones in open labs was established in Antwerp harbor in January 2019. The Russian government has also announced that a new experimental legal framework will be applied to the city of Moscow for AI experimentation.

Given that these various initiatives to create regulatory sandboxes for AI-based innovation have only recently been introduced, it is difficult to make concrete judgments about what impacts have been made by the regulatory sandboxes. There are only limited empirical data from which to draw any conclusions as to the extent regulatory sandboxes have succeeded in creating innovation as expected. At the same time, we do not yet fully comprehend the scope of privacy violations or security risks that consumers may be subjected to by AI algorithms.

Regulatory Sandboxes for Data Governance in Smart Cities

Although empirical findings are still limited, we can identify a number of key challenges in designing and implementing regulatory sandboxes for AI-based innovation in real-life settings. These include: how to guarantee compliance with regulations for safety, health, environment, security, and privacy, and to what extent regulations can be modified; how to share responsibility between the public and private sectors when accidents or problems have occurred; and how to manage accessibility, sharing, ownership, and use of data. In particular, data governance is a critical

challenge in fully utilizing the approach of regulatory sandboxes for AI-based innovation in the context of smart cities.

Various sectors are undergoing significant transformations by introducing data-driven innovation in smart cities. In the energy sector, distributed energy systems with peer-to-peer exchange of energy have become possible through blockchain technology, with photovoltaics provided through Solar-as-a-Service (SaaS). Smart meters and IoT technologies are providing highly sophisticated services for energy, health, and security to buildings and houses. In transportation, connected, autonomous, sharing, and electrified (CASE) challenges are radically changing the technologies and systems in the sector, and Mobility-as-a-Service (MaaS) is being explored aggressively through alliances among key players across the globe. In the health sector, Software as a Medical Device (SaMD) is being explored, and the diagnosis of cancers based on image recognition is considered especially promising.

An essential approach to stimulating data-driven innovation in smart cities is to foster data collection and sharing. A vast amount of various kinds of data would be collected from energy systems, public transportation, individual vehicles, and buildings, and many benefits would be expected from using that data for different types of innovation. For example, while the data collected through smart meters on energy consumption in households would be useful for optimizing energy use, that data could also be used for providing other services such as home delivery services. The data could tell delivery operators when residents would be at home, allowing them to adjust when to visit the house (Ohsugi & Koshizuka, 2018). The same data could also be used to provide health and security services to the residents of the house.

An open data approach facilitates collaborative efforts among stakeholders to create innovation for smart cities. In comparison to the conventional model of open innovation, which focuses on bilateral collaboration between firms, open innovation 2.0 is a new mode of innovation based on integrated

collaboration through experimentation with a wide range of actors in different sectors (Curley & Salmelin, 2018). Open data initiatives are increasingly considered as defining elements of emerging smart cities, which can be characterized as open innovation economies enabled by the participation of city residents, civic society, software developers, and local small- and medium-sized enterprises (SMEs) (Ojo, Curry & Zeleti, 2015). A recent study which analyzed patent applications in smart cities across the globe suggests that smart city policies have a positive impact on the rate of innovation, particularly in the high-tech sector (Caragliu & Del Bo, 2019).

There are many issues that we need to consider when implementing open data in smart cities. These include the types of data collected, who owns and has access to the data, for what purposes can the data be used, how the data are managed, and what incentives are provided to encourage data sharing to stimulate innovation while addressing concerns about privacy and security in smart cities. Although laboratory-level attempts have been made to integrate various types of datasets and sources on research data scattered across organizations, the scope and amount of data collected and shared needs to be expanded to scale-up innovative initiatives for actual implementation in smart cities. The quality control, error monitoring, and cleaning of data, as well as interoperability between various data standards, must be maintained to secure reliability. Organizational and legal frameworks need to be established concerning the ownership and accessibility of data, and to protect privacy and sensitive data. At the same time, it is also essential to keep a balance between open and proprietary data (Organisation for Economic Co-operation and Development, 2015b).

The collection and use of an extensive range of data, in particular, raises societal concerns in developing smart cities. The case of Sidewalk Toronto – a smart city project initiated in Toronto's waterfront by Alphabet, the parent company of Google – illustrates the seriousness of the concerns among citizens. There are various benefits expected to be provided to the residents and workers in the area, such as

ubiquitous high-speed Internet, intelligent traffic lights, smart shades in public spaces, underground delivery robots, and smart energy grids (Knight, 2019). The smart city plan would generate large quantities of data that could be used to optimize and improve technologies and services. However, some citizen groups were very concerned about the management of the collected data, and the Canadian Civil Liberties Association sued the City of Toronto in an attempt to block the project. After extensive consultation with citizens and companies in the city, the Master Innovation and Development Plan (MIDP) for Toronto was released in June 2019 (Sidewalk Labs, 2019a). The new plan emphasized community engagement and understanding of local needs in response to the concerns raised about building smart cities that are capable of tracking their inhabitants in unprecedented detail. Despite these efforts, the smart city project was eventually terminated (Doctoroff, 2020).

In trying to establish appropriate systems of data governance, it is useful to classify various types of data available in smart cities. Urban data can be defined as including personal, non-personal, aggregate, and de-identified data collected and used in physical or community spaces where meaningful consent before collection and use is difficult to obtain (Sidewalk Labs, 2019b). Non-personal data does not identify an individual and can include other types of non-identifying data not concerning people, such as machine-generated data about weather and temperature, and data on maintenance needs for industrial equipment. Aggregate data is about people in the aggregate and not about a particular individual, and is useful for answering research questions about populations or groups of people. Aggregate counts of people in an office space, for example, can be used in combination with other data, such as weather data, to develop an energy-efficiency program. De-identified data concerns an individual that was identifiable when the data was collected but has subsequently been made non-identifiable. Third-party apps and services can use properly de-identified data for research purposes, such as comparing neighborhood energy usage across a city. Personal data is usually the subject of privacy laws and includes any information

that could be used to identify an individual or that is associated with an identifiable individual. Individuals typically share their personal data with governments and businesses when applying for a license, shopping, or ordering a delivery service.

Digital transparency can be enhanced by providing easy-to-understand language that clearly explains the nature of data and privacy implications of digital technologies to citizens in smart cities (Lu, 2019). Through digital transparency, people are able to understand how and why data is being collected and used in the public realm through a visual language. For example, one hexagon conveys the purpose of the technology; another shows the logo of the entity responsible for the technology; and a third contains a QR code that takes the individual to a digital channel where they can learn more. In situations where identifying information is collected, a privacy-related colored hexagon can also be displayed by combining the technology type (video, image, audio, or otherwise) with the way that identifiable information is used (yellow for identifiable and blue for de-identified before first use, among others). This kind of approach could facilitate citizens' understanding and engagement in smart city projects.

A key question is what would be an appropriate governance system for urban data to maximize the potential of data-driven innovation while minimizing risks to individuals and communities. One approach is to establish a data trust, which is defined as a legal structure that provides for independent stewardship of data (Hardinges, Wells, Blandford, Tennison & Scott, 2019). With data trusts, the organizations that collect and hold data permit an independent institution to make decisions about who has access to data under what conditions, how that data is used and shared and for what purposes, and who can benefit from it. An independent urban data trust would be able to manage urban data and make it publicly accessible by default if appropriately de-identified (Sidewalk Labs, 2019b). An accountable and transparent process for approving the use or collection of urban data would ensure that local companies, entrepreneurs, researchers, and civic organizations can use urban

data. These data would be kept by the data trust and not be sold, used for advertising, or shared without the residents' permission.

In Japan, the Super City Initiative was started in October 2018 in an attempt to respond to the challenge posed by the fourth industrial revolution involving AI and IoT (Secretariat for the Promotion of Regional Development, 2020). The initiative requires that projects go beyond demonstrating a single technology, such as autonomous vehicles in a specific field, and to integrate it with other advanced services, such as cashless transactions and once-only application for administrative procedures, to comprehensively address a societal issue in a city. It also emphasizes that projects should incorporate the views and perspectives of the people living there, not simply the ideas promoted by the developers and suppliers of technologies. The super city initiative provides a particular legal procedure for deregulation that is specifically designed to simultaneously support regulatory reforms in different fields in an integrated manner. The broad regulatory changes involved in building smart cities often require dealing with multiple government agencies. In such cases, a top-down approach is taken; if a municipality obtains approval for smart city plans from its residents, the prime minister in the central government can direct agencies to make exceptions to the relevant regulations as needed. In June 2020, Japan's parliament just passed the "super city" bill, and the government is expected to soon begin taking applications from municipalities, with approvals starting in the summer (Miki, 2020).

In a super city, a data linkage platform plays a crucial role in facilitating close coordination among various services as the operating system (OS) of the city (Secretariat for the Promotion of Regional Development, 2020). A data linkage platform would be developed by professional vendors and operated by local governments, whereas private service providers would offer various services. As long as the residents of the super city agree, it would also be possible for either public agencies or private enterprises to provide services and the platform, making consent by the residents particularly crucial in data governance. For

example, when there are two separate systems for making taxi reservations and doctors' appointments, a data linkage platform can optimize taxi dispatching and appointment scheduling by connecting the relevant data in the two systems. The data linkage platform does not necessarily need to maintain an extensive central database, as data can be stored in separate databases in a distributed way. The providers of digital data and services are required to make their application program interfaces (APIs) open to the public, so that any information system can be developed through the data linkage platform. The super city initiative provides the operator of the data linkage platform with a right to request national and local governments and private enterprises to provide necessary data.

Several issues need to be addressed concerning data governance in smart cities through regulatory sandboxes. For the use of sophisticated services available in smart cities, personal data will be required on various aspects of the residents' lives. In the case of introducing an app connecting taxi-hospital reservations, the data linkage platform would ask the national or local government for personal data on the address, health status, and level of care needed by the elderly. The provision of such data would require the consent of the person in question in accordance with the law. On the other hand, relevant laws might allow the provision of such data without the permission of the person if there is a particular reason, such as contributing to the public interest. As local governments, businesses, or regional councils would make decisions in such cases, clear, transparent, and inclusive procedures are necessary for relevant stakeholders.

Another issue is how to reach a consensus among residents in smart cities. As residents are expected to agree on what kind of city they would like, and which areas they would target, the process of building a consensus needs to be well-integrated into the planning process. Furthermore, the methodologies and procedures for consensus-building need to be specified and institutionalized in an open and inclusive manner. It is also essential to consider how

to protect the rights of those residents who do not want to participate in the data governance scheme of smart cities. Residents need to form a consensus on where the balance should be located between the convenience of the advanced services that rely upon personal data and the risk of the data being used without their consent.

At the same time, the openness and interoperability of data in smart cities needs to be secured. In smart cities, it is often challenging to provide a cross-sectoral service because, typically, data is independent for each field and organization. Reusing and deploying such services to other cities is also difficult because the data system is specialized for each city. Moreover, the cost and labor required for functional expansion in the conventional data system increases, and services cannot easily be expanded to a larger scale. The provision of various services will be improved through close linkage and coordination of data in other systems and cities. APIs play a particularly significant role in facilitating interoperability and data flow. The design process of APIs defines conventions of data exchanges that influence interactions among the stakeholders involved (Raetzsch, Pereira, Vestergaard & Brynskov, 2019). It is essential to make APIs open, secure, and transparent, so that various kinds of data and sophisticated services are connected efficiently and effectively.

Coordinated efforts to share experiences in regulatory sandboxes at the international level will help to foster openness and interoperability to promote data sharing and use for innovation and transparency, as well as trust in managing and governing data to address concerns about privacy and security. So far, no global policy framework has yet been established on how to govern data for smart cities (Russo, 2019). For example, there is no shared set of rules concerning how sensor data collected in public spaces, such as by traffic cameras, should be used. It is of critical importance to explore guidelines and principles for the development and deployment of emerging technologies for smart cities by sharing good practices. As an international initiative to address these challenges, the G20 Global

Smart Cities Alliance on Technology Governance was launched in October 2019. The initiative aims to establish global standards for data collection and use, foster greater transparency and public trust, and promote best practices in smart city governance (World Economic Forum, 2019). Working together with municipal, regional, and national governments, as well as private-sector partners and city residents, the alliance intends to co-design, pilot, and scale-up policy solutions to help cities responsibly implement data-driven innovation. Such an international initiative will contribute to developing a global policy framework for smart cities by examining key issues concerning data governance, including privacy, transparency, openness, and interoperability, based on experiences through regulatory sandboxes in different locations.

Conclusion

Data-driven innovation plays a crucial role in tackling sustainability challenges. As the development of AI is accelerated by deriving new and significant insights from the vast amount of data generated during the delivery of services every day, training and adaptation is key to creating data-driven innovation. The development of cyber-physical systems such as smart cities is facilitated through the ready availability of and accessibility to data, and its mutual exchange and sharing with stakeholders in different sectors. Hence the new mode of data-driven innovation requires open, dynamic interactions with stakeholders possessing and generating various kinds of data. Close cooperation and collaboration in regards to data is crucial in the innovation process, from the development of novel technologies to deployment through field experimentation and legitimation in society.

It is critical to establish a proper system to govern data-driven innovation in the context of accelerating technological progress and deepening interconnection and interdependence. The speed of technological change with AI is remarkably fast, and it is accompanied by a significant degree of uncertainty in terms of consequences and side effects. Various types of technologies are increasingly becoming

interconnected and interdependent through data exchange and sharing among multiple sectors in smart cities, such as energy, buildings, transportation, and health. These characteristics make it difficult to explain or understand the process of innovation, and contribute to giving rise to a widening gap between technological and institutional changes. AI-based innovation becomes robust by involving the stakeholders who will interact with the technology early in development, obtaining a deep understanding of their needs, expectations, values, and preferences, and testing ideas and prototypes with them throughout the entire process.

Specifically designating geographical areas or sectoral domains, in the form of regulatory sandboxes, can facilitate data-driven innovation by allowing experimental trials of novel technologies and systems that cannot currently operate under the existing regulations. They provide a limited form of regulatory waiver or flexibility for firms to test new products or business models with reduced regulatory requirements, while preserving certain safeguards to ensure appropriate consumer protection. The aim is to provide a symbiotic environment for innovators to test new technologies, and for regulators to understand their implications for industrial innovation and consumer protection. Regulatory sandboxes help to identify and better respond to regulatory breaches by enhancing flexibility and adjustment in regulations, which would be particularly relevant in highly regulated industries, such as the finance, energy, transport, and health sectors.

The approach of regulatory sandboxes will play an especially essential role in governing data-driven innovation in smart cities, which inevitably faces a difficult challenge of collecting, sharing, and using various kinds of data for innovation while addressing societal concerns about privacy and security. Regulatory sandboxes can relax or adjust some of the relevant regulations, so that these new technologies can be tested for actual adoption and use. How regulatory sandboxes are designed and implemented can be locally adjusted, based on the specificities of the economic and social conditions and contexts,

to maximize the effect of learning through trial and error. Various types of new promising technologies can be verified, adopted, and integrated, effectively improving technological performance, reliability, and integration, as well as contributing to cost reduction. As AI-based innovation involves rapid technological change, uncertain market developments, and diverse social norms, there are many economic, ethical, and legal issues comprised of various interests and preferences. Regulatory sandboxes need to be flexible to accommodate the uncertainties of innovation, and precise enough to impose society's preferences on emerging innovation, functioning as a nexus of top-down strategic planning and bottom-up entrepreneurial initiatives.

Emerging cases of regulatory sandboxes for smart cities show that data governance is critical to maximizing the potential of data-driven innovation while minimizing risks to individuals and communities. With data trusts, the organizations that collect and hold data permit an independent institution to make decisions about who has access to data under what conditions, how that data is used and shared and for what purposes, and who can benefit from it. Alternatively, a data linkage platform can facilitate close coordination between the various services provided and the data stored in a distributed manner, without maintaining an extensive central database. The operator of the data linkage platform would require a right to request national and local governments and private enterprises to provide necessary data. APIs-linking data and services need to be open to the public so that any information system can be developed through the data linkage platform.

It is critically important that the data governance systems of smart cities are open, transparent, and inclusive. While the provision of personal data would require the consent of the person in question, the relevant law might allow the provision of such data without the permission of the person if there is a particular reason, such as contributing to the public interest. As local governments, businesses, or regional councils would be expected to make a decision, clear, transparent, and inclusive procedures are necessary

for relevant stakeholders. The process of building a consensus among residents needs to be well-integrated into the planning of smart cities, with the methodologies and procedures for consensus-building specified and institutionalized in an open and inclusive manner. It is also essential to respect the rights of those residents who do not want to participate in the data governance scheme of smart cities. As APIs play a crucial role in facilitating interoperability and data flow in smart cities, open APIs will facilitate the efficient connection of various kinds of data and sophisticated services. International cooperation will be critically important to develop common policy frameworks and guidelines for facilitating open data flow while maintaining public trust among smart cities across the globe.

Policy Recommendations

Recommendation 1: New policy approaches are required to govern data-driven innovation in the context of accelerating technological progress and deepening interconnection and interdependence.

Recommendation 2: Regulatory sandboxes should be established to facilitate data-driven innovation by allowing experimental trials of novel technologies and systems that cannot currently operate under the existing regulations through specifically designating geographical areas or sectoral domains.

Recommendation 3: Stakeholders should be involved from the early stages of technological development in order to obtain a deep understanding of their needs, expectations, values, and preferences, and to test ideas and prototypes with them throughout the entire process.

Recommendation 4: Regulatory sandboxes should be designed and implemented by incorporating the specificities of local economic and social conditions and contexts to maximize the effect of learning through trial and error.

Recommendation 5: Regulatory sandboxes need to be flexible to accommodate the uncertainties of innovation, and precise enough to impose society's preferences on emerging innovation, functioning as a nexus of top-down strategic planning and bottom-up entrepreneurial initiatives.

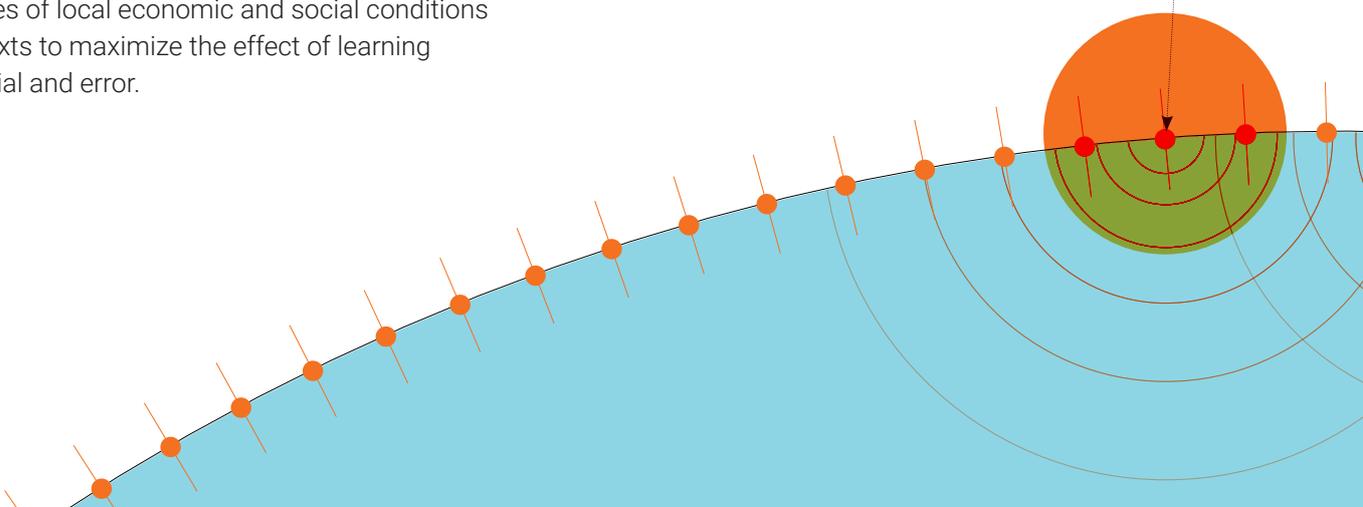
Recommendation 6: Data governance systems of smart cities should be open, transparent, and inclusive to facilitate data sharing and integration for data-driven innovation while addressing societal concerns about security and privacy.

Recommendation 7: The procedures for obtaining consent on the collection and management of personal data should be clear and transparent to relevant stakeholders with specific conditions for the use of such data for public purposes.

Recommendation 8: The process of building a consensus among residents should be well-integrated into the planning of smart cities, with the methodologies and procedures for consensus-building specified and institutionalized in an open and inclusive manner.

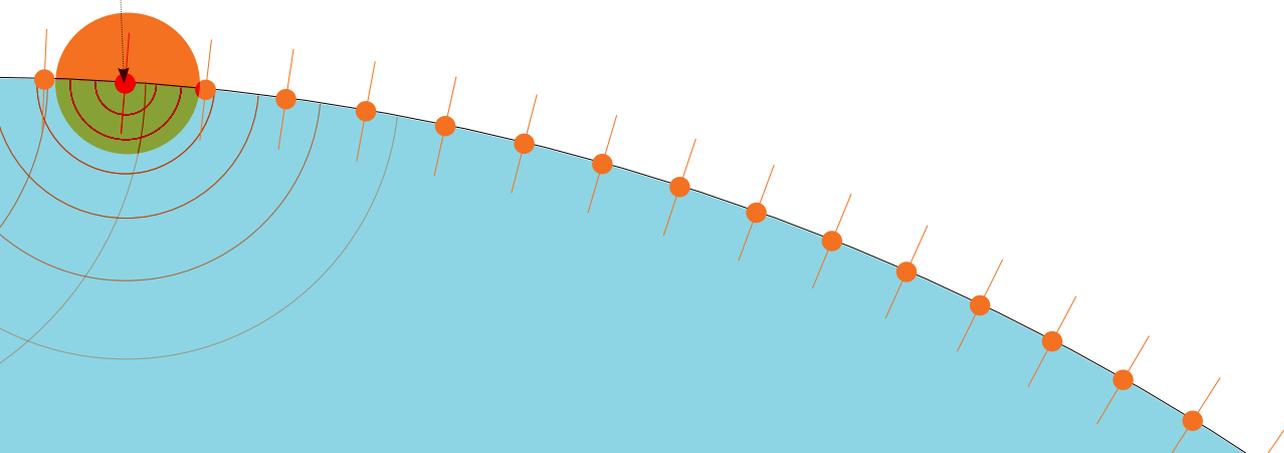
Recommendation 9: Application programming interfaces (APIs) should be open to facilitate interoperability and data flow for efficient connection of various kinds of data and sophisticated services in smart cities.

Recommendation 10: Common policy frameworks should be explored to develop guidelines for data collection and use, foster greater transparency and public trust, and promote interoperability and open data flow among smart cities across the globe.



References

- Allen, H. J. (2019). Regulatory Sandboxes. *George Washington Law Review*, 87(3), 579-645.
- Allen, H. J. (2020). Sandbox Boundaries. *Vanderbilt Journal of Entertainment & Technology Law, Forthcoming*, 22(2), 299-321.
- Beede, E. (2020, April 25). *Healthcare AI systems that put people at the center*. Retrieved from Google Blog: <https://www.blog.google/technology/health/healthcare-ai-systems-put-people-center/>
- Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020). A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-12.
- Boyd, A. (2018, May 9). *10 Drone Programs Get Federal OK To Break The Rules*. Retrieved from Nextgov: <https://www.nextgov.com/emerging-tech/2018/05/10-drone-programs-get-federal-ok-break-rules/148098/>
- Caragliu, A., & Bo, C. F. (2019). Smart innovative cities: The impact of Smart City policies on urban innovation. *Technological Forecasting and Social Change*, 142, 373-383.
- Care Quality Commission. (2020, March). Using machine learning in diagnostic services: A report with recommendations from CQC's regulatory sandbox. *Care Quality Commission*.
- Curley, M. (2016). Twelve principles for open innovation 2.0. *Nature*.
- Curley, M., & Salmelin, B. (2018). Data-Driven Innovation. In *Open Innovation 2.0: The New Mode of Digital Innovation for Prosperity and Sustainability*. Cham: Springer International Publishing.
- Doctoroff, D. L. (2020, May 7). *Why we're no longer pursuing the Quayside project – and what's next for Sidewalk Labs*. Retrieved from Medium: <https://medium.com/sidewalk-talk/why-were-no-longer-pursuing-the-quayside-project-and-what-s-next-for-sidewalk-labs-9a61de3fee3a>
- Energy Market Authority. (2017, October 23). Launch of Regulatory Sandbox to Encourage Energy Sector Innovations. *EMA*.



Federal Aviation Administration. (2019, December 10). *UAS Integration Pilot Program*. Retrieved from United States Department of Transportation: https://www.faa.gov/uas/programs_partnerships/integration_pilot_program/

Federal Ministry for Economic Affairs and Energy. (2019, July). Making Space for Innovation: The handbook for regulatory sandboxes. *BMW*.

Financial Conduct Authority. (2015). Regulatory Sandbox. *FCA*.

Financial Conduct Authority. (2017, October). Regulatory Sandbox Lessons Learned Report. *FCA*.

Fleischer, V. (2010). Regulatory Arbitrage. *Texas Law Review*, 89(2), 227-289.

Food and Drug Administration. (2019, April). Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. *FDA*.

Guston, D. H. (2014). Understanding 'anticipatory governance'. *Social Studies of Science*, 44(2), 218-242.

Hardinges, J., Wells, P., Blandford, A., Tennison, J., & Scott, A. (2019, April). Data trusts: lessons from three pilots. *Open Data Institute*.

Iizuka, M., & Ikeda, Y. (2019). "Regulation and innovation under Industry 4.0: Case of medical/healthcare robot, HAL by Cyberdyne." Working Paper Series #2019-038, Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT).

Ikeda, Y., & Iizuka, M. (2019, October). "International Rule Strategies for Implementing Innovation in Society: A Case Study of the Medical Healthcare Robot HAL." RIETI Policy Discussion Paper Series 19-P-016, Research Institute of Economy, Trade and Industry.

Information Commissioner's Office. (2019). ICO opens Sandbox beta phase to enhance data protection and support innovation. *ICO*.

Knight, W. (2019). Alphabet's smart city will track citizens, but promises to protect their data. *MIT Technology*.

Lu, J. (2019, April 19). *How can we bring transparency to urban tech? These icons are a first step*. Retrieved from Medium: <https://medium.com/sidewalk-talk/how-can-we-make-urban-tech-transparent-these-icons-are-a-first-step-f03f237f8ff0>

Miki, R. (2020, May 13). *Coronavirus pushes Japan closer to high-tech 'super cities'*. Retrieved from Nikkei Asian Review: <https://asia.nikkei.com/Politics/Coronavirus-pushes-Japan-closer-to-high-tech-super-cities>

Norwegian Ministry of Local Government and Modernisation. (2020). National Strategy for Artificial Intelligence. *H-2458 EN*.

OECD. (2015a). Data-Driven Innovation: Big Data for Growth and Well-Being. *OECD Publishing*.

OECD. (2015b). Making Open Science a Reality. *OECD*.

OECD. (2018). OECD Science, Technology and Innovation Outlook 2018.

OECD. (2019). Digital Innovation: Seizing Policy Opportunities. *OECD*.

Office of Gas and Electricity Markets. (2018a). Enabling trials through the regulatory sandbox. *ofgem*.

Office of Gas and Electricity Markets. (2018b). Insights from running the regulatory sandbox. *ofgem*.

Ohsugi, S., & Koshizuka, N. (2018). Delivery Route Optimization Through Occupancy Prediction from Electricity Usage. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, 842-849.

Ojo, A., Curry, E., & Sanaz-Ahmadi, F. (2015). A Tale of Open Data Innovations in Five Smart Cities. *2015 48th Annual Hawaii International Conference on System Sciences (HICSS-48)*, 2326-2335.

Pagallo, U., Casanovas, P., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., . . . Valcke, P. (2019). On Good AI Governance: 14 Priority Actions, A S.M.A.R.T. Model of Governance, and a Regulatory Toolbox. *AI4People*.

Pollman, E. (2019). Tech, Regulatory Arbitrage, and Limits. *European Business Organization Law Review*, 20(3), 567-590.

Raetzsch, C., Pereira, G., Vestergaard, L. S., & Brynskov, M. (2019). Weaving seams with data: Conceptualizing City APIs as elements of infrastructures. *SAGE Journals*, 6(1).

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Larochel. (2019). Machine behaviour. *Nature*, 568(7753), 477-486.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211-252.

Russo, A. (2019). World Economic Forum to Lead G20 Smart Cities Alliance on Technology Governance. *World Economic Forum*.

Secretariat for the Promotion of Regional Development. (2019). *The National Strategic Special Zones*. Retrieved from Cabinet Office, Prime Minister's Office of Japan: https://www.kantei.go.jp/jp/singi/tiiki/kokusentoc/supercity/supercityforum2019/supercityforum2019_EnglishVer.html

Secretariat for the Promotion of Regional Development. (2020). About the Super City Initiative. *Cabinet Office, Prime Minister's Office of Japan*.

Sidewalk Labs. (2019a). Sidewalk Labs Publishes Comprehensive Blueprint for the Neighbourhood of the Future. *Sidewalk Labs*.

Sidewalk Labs. (2019b). Toronto Tomorrow: A new approach for inclusive growth, Volume 2. *Sidewalk Labs*.

Stilgoea, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580.

Taeihagh, A., & Lim, H. S. (2019). Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, 39(1), 103-128.

World Economic Forum. (2019). Forum-led G20 Smart Cities Alliance will create the first global framework for smart city governance. *World Economic Forum*.

Yarime, M. (2017). Facilitating data-intensive approaches to innovation for sustainability: opportunities and challenges in building smart cities. *Sustainability Science*, 12(6), 881-885.

Zetzsche, D. A., Buckley, R. P., Arner, D. W., & Barberis, J. N. (2017). Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. *Fordham Journal of Corporate and Financial Law*, 23(1), 31-103.