

Response to “Enforcement Mechanisms for Responsible #AIforAll”

BY Tandem Research

January 15, 2021



Authors: Amrita Vasudevan, Angelina Chamuah, Urvashi Aneja

Table of Contents

PART I: RISK-BASED APPROACHES TO AI GOVERNANCE

1.1 Brief overview of Risk-Based Regulations with respect to AI

1.2 Risk-Based Regulations: Benefits and Rationale for Adoption

1.3 Challenges of Risk-Based Regulations for AI

1.3.1 Risk-based approaches entail choosing between competing priorities and values

1.3.2 Risk-based approaches fail to account for distributional justice

1.3.3 Reliance on risk-based analysis can also lead to excessive risk-taking and provide a false illusion of control

1.3.4 Risk-based approaches divert attention away from societal risks towards institutional risks.

1.4 Other Approaches and Measures for AI Governance.

1.4.1 Precautionary Principle

1.4.2 Human Rights Approach

1.4.3. Algorithmic Impact Assessments

1.4.4 Product Liability Frameworks

PART II: EVALUATING THE COUNCIL FOR ETHICS AND TECHNOLOGY

2.1 Understanding the Nature and Functions of Regulatory Oversight Bodies

2.2 Unpacking the Relationship between the Oversight Body and Sectoral Regulators

2.3 Instituting Measures for Independence, Transparency and Accountability

2.4 Unpacking the use of standards and self -regulation by the private actors

PART III: MOVING TOWARDS ANTICIPATORY GOVERNANCE AND RESPONSIBLE RESEARCH AND INNOVATION

3.1 Introduction to Responsible Research and Innovation

3.2 Components of RRI

3.2.1 Product Dimension in the Indian context

3.2.2 Process Dimension in the Indian context

EXECUTIVE SUMMARY

The NITI Aayog working document on “Enforcement Mechanisms for Responsible #AIforAll”¹ advocates a flexible and context-specific, risk-based approach to regulating AI which is to be effectuated by an oversight body. The document argues that a one-size-fits-all approach is not sustainable in the light of the rapid growth of AI in the past few years, across diverse industries and sectors in India.

The working document suggests that the oversight body would work in an advisory capacity with sectoral regulators, since existing regulatory mechanisms are best placed to enforce rules. The functions of the oversight body would include monitoring and updating the principles of responsible AI (identified by the NITI Aayog in a previous working document “Towards responsible #AIforAll”)²; clarifying what responsible behaviour entails (by leveraging existing standards and developing guidelines); and co-ordinating with sectoral regulators (including helping them identify risks with the use of AI). While an ‘ethical committee’ will review the use of AI by public authorities, private sector actors will be allowed to self-regulate, except in cases deemed high-risk, where adherence mechanisms will be mandated by the oversight body in consultation with sectoral regulators.

The proposed oversight advisory body will be formed as a multidisciplinary ‘Council for Ethics and Technology’. The council will consist of experts from different fields as well as representatives from industry, standard setting bodies and civil society.

¹NITI Aayog. (2020). (rep.). *Working Document: Enforcement Mechanisms for Responsible #AIforAll*. Retrieved from <https://niti.gov.in/sites/default/files/2020-11/Towards-Responsible-AI-Enforcement-of-Principles.pdf>

²NITI Aayog. (2020). (rep.). *Working Document: Towards Responsible #AIforAll*. Retrieved from https://niti.gov.in/sites/default/files/2020-11/Towards_Responsible_AIforAll_Part1.pdf

Our response to the NITI Aayog's working document "Enforcement Mechanisms for Responsible #AIforAll" is organised in three parts: part I examines risk-based and principle based approaches to the regulation of AI, and considers these with regard to alternative and complementary frameworks and approaches; part II examines the role of the oversight body; and part III focuses on the need for upstream management of technological innovation and the role of responsible innovation with regard to AI.

The adoption of risk-based approaches for regulating AI has found purchase across a number of jurisdictions globally. Risk-based approaches allow for flexibility and economize regulatory attention to cases deemed as high-risk with a high probability of risks occurring. However they can also create regulatory blind spots with regard to disparate impacts on vulnerable populations and systemic risks. In part I, **we argue that risk-based approaches must not be viewed as neutral policy instruments. Instead, they should be seen as a complex set of choices regarding which risks will be prioritised, and the degree of risks that will be tolerated. These choices are grounded in values and cannot be resolved through objective assessments alone.**

Risk-based regulatory approaches to AI also face methodological and epistemic challenges. For instance, **not all AI risks may be amenable to categorisation of low, medium and high thresholds. Though some risks of AI may have a low impact, their cumulative effect could be overwhelming.** Secondly, **risk-based approaches based on the principle of welfare maximisation are not equipped to safeguard against the disproportionate impact of AI harm on minorities and marginalised populations.**

Therefore, for a risk-based approach to be effective, **regulators must be explicit about the criteria of selection of the risks** to be regulated as well as the risk appetites. Furthermore, the **selection of risks must make room for open and transparent public deliberation**, as even the lowest relative risk may not be equal to acceptable risk for all stakeholders. Secondly, any form of risk-based calculation should **prioritise and uphold constitutionally guaranteed rights and liberties, and place greater weight on the disproportionate impact of AI on vulnerable populations.**

We also note that while risk-based approaches can remedy ham-fisted regulatory interventions and promote innovation, they **must be combined with complementary approaches which can address some of the more egregious harm and systemic impacts of AI.** Several existing legal frameworks can also be extended to address the governance of AI alongside risk.

For example, India has developed a **jurisprudence around the precautionary principle** in the environmental sector, which says that scientific evidence is not necessary for regulatory intervention where there is a threat of serious and irreversible damage. **Under the discussion of the precautionary principle we note that the principle's malleability is well suited to address the unpredictability of the risks of AI as well.**

Similarly, a **product liability framework, which we see as one part of AI governance, could also function alongside a risk-based regime.** Fault-based liability is currently not totally appropriate for AI products and services. This could be remedied with suitable tweaks and amendments. We suggest that India, like the EU, should study the failings of the Consumer Protection Act with regard to AI products and services, and suggest how it could be remedied.

Finally, the **evolving frameworks for AI impacts assessments can provide much needed oversight in the process of AI deployment**, and can help identify points of regulatory intervention through multiple iterative cycles of AI development and deployment.

Part II examines the origins of regulatory oversight bodies and compares NITI Aayog's proposed institutional framework with those of other similar oversight bodies. **We argue that India requires a strong regulatory body to steer the regulation of AI (as well as other emerging tech), as the present regulatory landscape has critical lacunae.** We recommend that NITI Aayog should reconsider the proposal that the oversight body is to perform only an advisory function. **We also point out that the working document fails to provide key information on the location of the oversight body, information that one would need to understand the extent of the body's influence as well as its independence. We also urge NITI Aayog to include mechanisms to secure measures for transparency, accountability and consultation which are currently lacking.**

Lastly, part III makes a case for **upstream governance of AI**. The Responsible Research and Innovation (RRI) framework is a useful way of thinking through non-consequentialist framings of responsibility. RRI has been interpreted by scholars to include product and process dimensions. Product dimension explores how the innovation process can be engineered to not only mitigate risks to locally evolved values (such as the principles for responsible AI) but actively promote them. The process dimension seeks to incorporate public engagement suitably early in the innovation process to have meaningful impact. Both these dimensions, we posit, should find resonance within the regulatory landscape of emerging tech and the functions of the oversight body. **We agree with the sentiment of RRI that rolling back of technology is next to impossible and so anticipatory governance focused upstream is necessary.**

We applaud Niti in producing this document and initiating this conversation around Responsible AI. To close, we would like to draw attention to a few additional points motivated by the NITI Aayog document on 'Towards Responsible AI for All'.

- Responsible AI is not only about beneficial applications, or reducing harmful effects of existing applications, but first order questions about where AI should be applied at all. A responsible AI strategy must thus include considerations of where to apply AI, not just how to apply it responsibly.
- Do not promote the indiscriminate uptake of artificial intelligence. Strategy should not be adoption driven, but problem driven. AI solutions to problems must be evaluated against non-AI approaches.
- It is essential to place public interest, the protection of individual rights, non-discrimination, and equal access to resources and participation at the core of any strategy on AI.
- Adoption challenges are not the same as harms and risks. The framing of challenges as harms, gives off the false notion that once the challenges have been overcome through quick fixes and technical solutions, industry or government should be able to move ahead with adoption. In our understanding, challenges pertain to roadblocks in adoption - such as infrastructural gaps, lack of human resource and institutional gaps in building, developing and deploying AI. What the document identifies as challenges pertain to deep seated societal risks immanent in the nature of AI use today.
- Current legal and technical solutions are likely to be inadequate to govern AI. Current legal frameworks do not provide adequate protection for individuals. The Personal Data Protection Bill for example does not take into consideration the issues related inferential data. The focus on individual data protection is also inadequate for AI systems which are based on aggregated data. Equally, technological techniques are likely to outcompete one another - for example, while anonymization was seen as a way to protect individual privacy, new advances in ML are already rendering it obsolete.

- We need to place moratoriums on potentially dangerous use cases such as the use of facial recognition, for instance, the use of facial recognition for behaviour profiling, surveillance and discrimination.
- There should be mandatory disclosure of any algorithmic decision making systems used in the public sector. The self assessment guide merely makes a mention of public auditing of AI systems but does not consider the significance in active disclosure around its deployment. In achieving transparency and accountability through the use of assessment frameworks, the results of assessments should also be made public through the establishment of public registers. Mandatory disclosure algorithmic decision making systems used in the public sector and algorithmic impact assessments are vital to the functioning of a health democracy and ensuring transparency and accountability.

PART I: RISK-BASED APPROACHES TO AI GOVERNANCE

1.1 Brief overview of Risk-Based Regulations with respect to AI

The governance of AI is an emerging issue in many jurisdictions globally, wherein regulation is considered necessary both to encourage innovation and manage associated risks. Aside from the promulgation of high level principles, both by supra-national institutions as well as private consortiums, such as the OECD and the Partnership for AI (PAI), jurisdictions such the EU and U.S. have also proposed risk-based regulations for AI governance.

For instance, in February 2020 the European Union published its White Paper on Artificial Intelligence - A European approach to excellence and trust. The white paper outlines the EU's approach for a regulatory framework for AI. In its proposed approach, the Commission differentiates between 'high-risk' and 'non-high-risk' AI applications. Only high-risk applications should be in the scope of a future EU regulatory framework.³ Similarly, in January 2020, the U.S. government also released draft rules or the “Guidance for Regulation of Artificial Intelligence Applications” which advocates for a risk-based approach to AI regulations. The document states that while some risks are acceptable, AI applications which pose “unacceptable harm, or harm with greater harm than benefits” can be regulated.⁴

Thus, NITI Aayog’s proposal for the adoption of a flexible risk-based approach must also be seen as part of a wider trend amongst regulators to adopt the approach. While risk-based approaches (RBA) are increasingly applied to AI governance, the

³ European Commission. (2020). White paper on artificial intelligence—a European approach to excellence and trust.

⁴ Office of Management and Budget, U.S.A (2020). Re: Guidance for Regulation of Artificial Intelligence Applications.

approach itself is not new. Models of risk based regulations are spread across diverse sectors and jurisdictions in the world. For instance, risk-based regulatory approaches have been widely adopted in the financial sector in jurisdictions such as the U.S., U.K., Canada and Australia⁵. Similarly, risk-based regulations have a long history in several other domains, such as environment and food safety⁶, healthcare⁷ and education⁸. The application of the approach in different sectors provides important lessons highlighting the benefits and challenges of the approach. In this section, we examine the contours of risk-based approaches and it's benefits and challenges with respect to AI governance.

1.2 Risk-Based Regulations: Benefits and Rationale for Adoption

A risk-based regulatory approach can be defined as a “commitment to a sophisticated form of prioritization resulting in an approach to applying regulation proportionately to the risk posed by the activities of an organization”.⁹ In it's idealised form, Julia Black, professor of Law at the London School of Economics argues, risk-based regulation offers an ‘evidence-based means of targeting the use of resources and of prioritizing attention to the highest risks in accordance with a transparent, systematic, and defensible framework’. While these definitions provide a working understanding of the approach, in reality, risk-based regulations embrace a broad range of approaches and techniques. As noted by sociologist Bridget M. Hutter, it can refer to an ‘entire perspective of governance, while also referring to the piecemeal adoption of risk based tools and an uneven use of the

⁵ Black, J. (2004). The development of risk based regulation in financial services: Canada, the UK and Australia. *ESRC Centre for the Analysis of Risk and Regulation Research Report*.

⁶ Black, J. (2005). The emergence of risk-based regulation and the new public risk management in the United Kingdom. *Public law*, 512.

⁷ Beaussier, A. L., Demeritt, D., Griffiths, A., & Rothstein, H. (2015). Why risk-based regulation of healthcare quality in the NHS cannot succeed.

⁸ Simkovic, M. (2013). Risk-based student loans. *Wash. & Lee L. Rev.*, 70, 527.

⁹ Adil, M. (2008). Risk-based regulatory system and its effective use in health and social care. *The journal of the Royal Society for the Promotion of Health*, 128(4), 196-201.

language and rhetoric of risk'.¹⁰ However, at the very minimum, **a risk-based approach involves the adoption of a technical risk-based tools, such as cost-benefit approaches, and risk assessment techniques¹¹ to measure, quantify and introduce levels of hierarchy (such as high, medium and low) between diverse sets of risks.**

Risk-based approaches are often contrasted with the rules-based approaches. In fact, as the literature on the historical dimensions of risk-based approaches point out, the development of risk-based regulations (as well the concomitant development of arms-length regulatory bodies) have emerged in direct opposition to the rules-based frameworks followed before, particularly in advanced economies such as the U.K and U.S. in the 1980s.¹² By focussing on risks rather than rules (which may be difficult to apply in every given situation), a risk based approach allows regulators to be selective. According to Black, 'selection' itself is not unique or new, but by relying on cost-benefit analysis, risk-based frameworks allow this selection to be made explicit and provide a framework of analysis in which selections can be made.¹³

The benefits of the move away from a rules-based approach to a more flexible risks based approaches has been validated by some research. For instance, in recent times, research conducted by the Centre for European Policy Studies on the multi-faceted relationship between (EU) regulation and innovation in the economy, has found that 'regulation can (indeed) at times be a powerful stimulus to innovation.. (However) more prescriptive regulation tends to hamper innovative activity, whereas the more flexible regulation is, the better innovation can be stimulated. Lower compliance and red-tape burdens have a positive effect on innovation.'¹⁴ Similarly, in the environmental sector, movement away from rules to a more flexible risk-

¹⁰ Hutter, B. M. (2005). *The attractions of risk-based regulation: accounting for the emergence of risk ideas in regulation* (Vol. 33). London: CARR.

¹¹ *ibid*

¹² Hutter, B. M. (2005). *The attractions of risk-based regulation: accounting for the emergence of risk ideas in regulation* (Vol. 33). London: CARR.

¹³ Black, J. (2010). Risk-based regulation: Choices, practices and lessons being learnt.

¹⁴ Pelkmans, J., & Renda, A. (2014). Does EU regulation hinder or stimulate innovation?.

based approach created incentives towards the development of technical risk assessment capabilities, amongst contaminated site owners.¹⁵

Aside from providing flexibility, Hutter and Black, have identified several reasons why governments and regulatory institutions have taken to risk-based regulations (RBA). One perceived benefit is that RBA can economize regulatory burden and make better use of limited funds by targeting activities that pose the highest risk and in-turn can lower burdens for a variety of lower-risk sectors and firms. Lowering burdens can then improve compliance and allow firms to benefit from a more level playing field. Second, is the perceived objectivity of the cost-benefit analysis model. According to Black, risk based tools came to be seen as efficient instruments for making policy choices and aiding decision-making. They are well regarded as particularly helpful in resolving any 'conflict' between differing interests groups when determining appropriate levels of risk management. Their apparent objectivity and transparency could be used to explain the allocation of resources, in a way which was well tested and trusted by the business community.¹⁶

A third, more implicit reasoning behind adopting risk-based approaches is to maintain and convey legitimacy. As Black notes, the rhetoric of 'risk management' and 'risk-based' approaches combine a sense of strategy and control which can be politically compelling.¹⁷ Similarly, Hutter also argues that RBA enables regulators to demonstrate objectivity, and objectively demonstrate the steps taken to calculate and manage risks.¹⁸

¹⁵ Rothstein, H., Irving, P., Walden, T., & Yearsley, R. (2006). The risks of risk-based regulation: Insights from the environmental policy domain. *Environment international*, 32(8), 1056-1065.

¹⁶ Supra note 13

¹⁷ Supra note 6

¹⁸ Supra note 12

However, risk-based approaches are not without challenges. Despite its widespread endorsement, the application of the risk-based approach has also led to regulatory failures. One of the most prominent examples is failure of risk-based regulations is the 2007–2009 financial crisis, which was as partly a result of implementation challenges as well as an excessive regulatory focus on individual institution risk rather than systemic risk.¹⁹ Challenges to risk-based approaches come from both intrinsic and extrinsic factors, which not only have to do with the categorical lens of risk, but also the institutional and regulatory contexts in which these are mired.

1.3 Challenges of Risk-Based Regulations for AI

Risk-based approaches may provide a valuable entry point for ascertaining a regime of flexible and differentiated regulations where not all AI applications or sectors are regulated alike. However, **the adoption of such a framework would still require safeguards that can protect vulnerable communities and groups. Similarly, the approach needs safeguards to ensure transparency and accountability in determining what are acceptable risks and which risks to regulate and ensure that institutional risks are not prioritized over societal risks.**

¹⁹ Acharya, V. V., Cooley, T. F., Richardson, M. P., & Walter, I. (2011). Market failures and regulatory failures: Lessons from past and present financial crises.

Some of the key challenges and issues related to risk-based approaches, particularly with respect to AI, are:

1.3.1 Risk-based approaches entail choosing between competing priorities and values.

A cross-country analysis of risk-based approaches shows that not every country/geography adopts the approach evenly, or in the same way. While there are commonalities to adoption of risk-based frameworks, much depends on the ‘fit’ of the approach to the character of the polity.²⁰ Each aspect of a risk-based framework involves a complex set of choices. In the elaborating the commonalities between different risk-based frameworks, Black notes that, **aside from determining the question of which risks to regulate (i.e. risk to society, risk to minorities, risks to the economy, environmental risks) and where it will prioritize its attention and where it will not, regulators also have to determine their own risk appetite—what type of risks is it prepared to tolerate and at what level?**²¹

Thus, Black notes, **‘at the heart of risk regulation is the very difficult issue of determining what is an acceptable risk or acceptable cost?’ These determinations are often driven by politics and power rather than evidence alone.**²² For example, Jasanoff notes that during the GMO debates, both the U.S and E.U took widely different approaches to regulation. While genetically modified foods were considered safe, and a non-issue by U.S regulators, E.U regulators took the opposite stance. According to Jasanoff, the trustworthiness of science in each policy context flows

²⁰ Rothstein, H., Borraz, O., & Huber, M. (2013). Risk and the limits of governance: Exploring varied patterns of risk-based governance across Europe. *Regulation & Governance*, 7(2), 215-235.

²¹ Black, J., & Baldwin, R. (2010). Really responsive risk-based regulation. *Law & policy*, 32(2), 181-213.

²² Supra note 13

not so much from science's transcendental claim to delivering the truth as from the production and deployment of scientific knowledge within particular traditions of garnering credibility and maintaining trust.²³

Hutter also notes that the institutional landscape of regulation changed with the adoption of risk-based approaches. From sector based regulation we began to witness more domain specific regulation that required umbrella agencies that act as super-regulator. **The difficulty with coordinating activities across agencies from a singular standpoint of risk is that different agencies may employ different regulatory approaches that may not complement each other ; some may even oppose a risk based approach**²⁴.

1.2.2 Risk-based approaches fail to account for distributional justice.

Adjudicating between competing priorities in the context of AI becomes more problematic due to AI's disparate impacts. **The risks of AI are differentially distributed and there is substantial evidence that ML algorithms impact minorities, women, and other protected classes disproportionately.** For instance, facial recognition programs are more likely to misidentify people with darker skin²⁵; using AI for assessing employability can provide lower scores to women who graduated from women's colleges or who participated in women's sports.²⁶ Minority

²³ Sheila Jasanoff (2002) Citizens At Risk: Cultures of Modernity in the US and EU, *Science as Culture*, 11:3, 363-380

²⁴ Supra note 12

²⁵ Simonite, T. (2019, July). The Best Algorithms Still Struggle to Recognize Black Faces. Retrieved January 15, 2021, from <https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>

²⁶ Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved January 15, 2021, from <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-id>

groups have been shown to be falsely flagged as being more likely to be recidivists than whites.²⁷ Not all AI risks may be amenable to low, medium and high thresholds. Even though some risks of AI may have low impacts, their cumulative effects could still be overwhelming.

Along with disparate impacts, the technological gains from AI may also not be distributed equally. For instance, capital and efficiency gains from the use of AI in enterprises may not trickle down to employees. However, as Eidenmüller notes, welfare maximization has become a mainstream goal of AI regulation.²⁸ The problem with cost-benefit analysis and the utilitarian logic of welfare maximization is that welfare is difficult to calculate. **In maximizing welfare and minimizing the risks, a purely utilitarian logic is likely to give less weight to risks borne by minorities, and marginalised populations, who might however need it the most.** Further, the utilitarian logic of welfare does not engage with distributional justice and is narrowly concerned with the maximization of welfare (of some). In the context of AI, optimising for maximum efficiency gains using a utilitarian lens for instance does not indicate how those efficiency gains will be redistributed and therefore the outcomes of differential risks. Even a consideration of human rights is contingent on whether they maximise welfare or not; human rights within this framework do not hold value as an independent goal.²⁹

²⁷ Larson, J., Angwin, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias. Retrieved January 15, 2021, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

²⁸ Eidenmüller, H. (2019). Machine Performance and Human Failure: How Shall We Regulate Autonomous Machines. *J. Bus. & Tech. L.*, 15, 109.

²⁹ *ibid*

1.2.3 Reliance on risk-based analysis can also lead to excessive risk-taking and provide a false illusion of control.

Risk-based analysis centres around the quantification of risks.³⁰ Quantification implies that risk can be predicted and controlled. However, in the research on the failure of the Taurus project in the U.K, H Drummond, Professor of Decision Sciences, University of Liverpool notes that, 'risk analysis and other managerial techniques, instead of reducing uncertainty, may actually facilitate escalation (of risks) by fostering an illusion of control.'³¹ Similarly, going back to the 2008 financial crisis example, legal scholar Julie Cohen, notes that reliance on risk assessment and risk management discourses can induce unwarranted complacency, and also can encourage excessive risk-taking.³² Prior to 2008, banks became too confident in their analysis of risks and strategies of risk-management, which then led to excessive risk taking, ultimately resulting in the crash.

According to Rothstein et. al., risk-based regulations can be expected to face epistemic challenges. Regulation often asks questions of science, which science is not in a sufficiently advanced state to answer. Such issues create considerable scope for regulatory uncertainty and conflict.³³ While this has been witnessed in the context of other sectors, with AI such epistemic challenges are likely to be felt even more. Artificial intelligence as a field is constantly evolving - the discovery of new AI risks have accompanied the growth of new techniques and applications of AI.

³⁰ Drummond, H. (1996). The politics of risk: trials and tribulations of the Taurus project. *Journal of Information Technology*, 11(4), 347-357.

³¹ *ibid*

³² Cohen, J. E. (2016). The regulatory state in the information age. *Theoretical Inquiries in Law*, 17(2), 369-414.

³³ Rothstein, H., Irving, P., Walden, T., & Yearsley, R. (2006). The risks of risk-based regulation: Insights from the environmental policy domain. *Environment international*, 32(8), 1056-1065.

While risk-based regulatory measures can be more forward-looking and anticipatory than rules-based approaches, the inability to identify unexpected risks of AI will also leave very little provisions in place to protect against unforeseen harms. Using risk-based logic to make decisions at the edge of scientific understanding may impart a false sense of security, masking uncertainties that may then be overlooked.

1.2.4 Risk-based approaches divert attention away from societal risks towards institutional risks.

Research shows that regulators are not simply optimising for societal risks (such as threat to privacy, discrimination, inequality etc.), but the adoption of risk based approaches are in fact also aimed at reducing institutional risks.³⁴ According to Rothstein et. al. enforcement failures, liabilities and damage to reputation can create problems for regulatory decision-making and even threaten the legitimacy of regulatory organisations themselves. Thus, risk-based regulation also accounts for risk to the regulatory institution, in addition to societal risks.³⁵

In doing so, the approach may also have the unintended consequence of focusing policy attention on problems that carry high institutional risks for the regulator at the expense of those that carry high risks to society. They give the example of train accidents versus road accidents. Even though statistically car accidents result in far more deaths than trainwrecks, public outcry and political fallout from train accidents are far greater. Therefore, regulators often pay greater attention to the latter than the former. The adoption of the activity-based costing and management

³⁴ ibid

³⁵ Rothstein, H., Irving, P., Walden, T., & Yearsley, R. (2006). The risks of risk-based regulation: Insights from the environmental policy domain. *Environment international*, 32(8), 1056-1065.

system (also known as ABC/M) in environmental regulation is also a case in point - regulators throughout the world adopted it because it had a defensible basis for their actions and implementation was straightforward.³⁶

Further, introduction of flexibility into the system also creates its own challenges. For instance, the introduction of flexibility with regard to environment protection and chemical contamination regulations led to endless debates about how clean is clean enough. Increased flexibility in the face of scientific uncertainty, thus, created challenges for the regulators charged with verifying how clean is clean.³⁷

With regard to risk-based regulations therefore we recommend that selection criteria for risks be made more explicit and clear. The selection of risks must incorporate open and transparent public deliberation processes, for even the lowest relative risk may not be equal to acceptable risk for all stakeholders.

We also recommend that NITI Aayog consider fortifying a risk-based approach with other complementary frameworks and legal procedures, which can help overcome regulatory blinspots with regard to systemic risks (for instance, impact on the environment), and expand the category of risks to include qualitative considerations of redistributive justice, human rights and constitutional principles. Ultimately, the purpose of AI governance should not simply be pegged to an accumulative logic of spurring innovation and economic growth, instead AI governance should aim to advance equitable and sustainable technological transitions.

³⁶ *ibid*

³⁷ Rothstein, H., Irving, P., Walden, T., & Yearsley, R. (2006). The risks of risk-based regulation: Insights from the environmental policy domain. *Environment international*, 32(8), 1056-1065.

1.4 Other Approaches and Measures for AI Governance

In recent years, both civil society and academia have forwarded multiple frameworks and perspectives for the governance of AI. Some of these frameworks, such as international human rights law, can be viewed as alternative approaches, whereas others such as the precautionary principle already sit within a risk-based approach. In this section, we provide a brief overview of some of the proposed frameworks for the governance of AI - such that elements of these approaches should be considered further by NITI Aayog as alternates and additions to the risk-based approach.

1.4.1 Precautionary Principle

The precautionary principle can be understood to be part of a risk based approach to regulation, but falling at the more risk-averse end of the spectrum. The application of the principle is most often discussed vis a vis environmental regulation, but holds important lessons for emerging tech. **AI regulation faces many of the same problems as the environmental sector particularly of unpredictability and consequently the lack of conclusive scientific evidence. The principle presents us with an opportunity to institute a 'stop and think' mechanism in the regulation of a sector where the slogan 'move fast and break things' has become unsustainable.** It allows regulators to intervene and discuss evidence even if the same is inconclusive, rather than waiting and watching for harm to occur.

In regulatory practice, there have been largely two ways of approaching uncertainties, either through cost-benefit analysis or precautionary principle. Under the former approach, as expounded by legal scholars like Cass Sunstein only

in case of potential catastrophes or worst-case scenarios where risk cannot be assigned and there are great uncertainties should a precautionary maximin rule be applied, and the worst case scenario eliminated.³⁸ Proponents of the precautionary principle on the other hand explains Douglas Kyser, Professor of Law at Yale Law School, “emphasize the limits of human knowledge and the frequency of unpleasant surprises from technology and industrial development; thus, they advocate an ex-ante governmental stance of precaution whenever a proposed activity meets some threshold possibility of causing severe harm to human health or the environment.”³⁹ The use of the precautionary principle, in the environmental sector, has been criticized for preventing firms from pursuing more innovative methods to limit the level of pollution. On the other hand, it has been argued that risk expresses “human life, endangered species, old-growth forests, and other protected subjects of environmental, health, and safety law” in terms of monetary value, and this begs serious ethical questions.⁴⁰ **The kinds of costs that the precautionary principle on the other hand accounts for are also more broad-based- it is concerned with the dynamic, inter-generational costs, rather than merely static welfare-maximization.**⁴¹

The Supreme Court of India, in *Vellore Citizen Forum v. Union of India* has held that, “Where there are threats of serious and irreversible damage, lack of scientific certainty should not be used as a reason for postponing measures to prevent environmental degradation” and further the “onus of proof is on the actor or developer or industrialist to show the

³⁸ Sunstein, C. R. (2020). Maximin. *Yale J. on Reg.*, 37, 940.

³⁹ Kysar, D. A. (2010). *Regulating from nowhere: Environmental law and the search for objectivity*. Yale University Press.

⁴⁰ *ibid*

⁴¹ Author does admit that interest groups that traditionally supported cost-benefit analysis (conservative, anti-regulatory groups) and those that opposed cost-benefit analysis (progressive groups) are both using cost-benefit analysis to buttress their views on regulation. Livermore, M. A., & Revesz, R. L. (2014). Environmental law and economics. *Oxford Handbook of Law and Economics*. Oxford University Press (Forthcoming).

actions are environmentally benign.”⁴² The National Green Tribunal has used the precautionary principles on the basis of scientific data on the risk to human health and the environment, although the same may have been inconclusive, tentative or in dispute.⁴³

AI can potentially cause widespread harm (including risk lives, fundamental rights, social structures and institutions etc.) either deliberately (in unethical ways) or accidentally (when systems behave in unintended ways). But there are also structural, cumulative risks from AI that are difficult to detect and thus predict.⁴⁴ This unforeseeability of AI must be highlighted.⁴⁵ **There are thus well justified reasons to apply the precautionary principle to risky applications of AI, even if the data is inconclusive, tentative or in dispute, when human lives, rights and the very fabric of society are at stake. Thus one way out of the policy paralysis induced by uncertainties of AI, is by using the precautionary principle.**⁴⁶ Once a technology is rolled out it becomes intractable and irrevocable due to a number of factors including the high cost of shifting to alternatives and the reluctance of users to adopt new technologies once they become accustomed to existing ones. Therefore, it is important to apply the precautionary principle suitably early in the causal chain.⁴⁷

⁴² Vellore Citizen’s Welfare Forum v Union of India (1996) 5 SCC 647 at 658 in itanjali Nain Gill, The Precautionary principle,its interpretation and application by the Indian judiciary: ‘When I use a word it means just what I choose it to mean-neither more nor less’ Humpty Dumpty. Environmental Law Review 2019, Vol. 21(4) 292–308

⁴³ Gitanjali Nain Gill, The Precautionary principle,its interpretation and application by the Indian judiciary: ‘When I use a word it means just what I choose it to mean-neither more nor less’ Humpty Dumpty. Environmental Law Review 2019, Vol. 21(4) 292–308

⁴⁴ Zwetsloot, R., & Dafoe, A. (2019, October 31). Thinking About Risks From AI: Accidents, Misuse and Structure. Retrieved January 15, 2021, from <https://www.lawfareblog.com/thinking-about-risks-ai-accidents-misuse-and-structure>

⁴⁵ Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, 29, 353.

⁴⁶ Kuziemski, M. (2018, May 01). A Precautionary Approach to Artificial Intelligence by Maciej Kuziemski. Retrieved January 15, 2021, from <https://www.project-syndicate.org/commentary/precautionary-principle-for-artificial-intelligence-by-maciej-kuziemski-2018-05?barrier=accesspaylog>

⁴⁷ Claudia Som,et al, The Precautionary Principle in the Information Society, Human and Ecological Risk Assessment, 10: 787–799, 2004

Scholars have been cautious of employing the precautionary principle in all cases. For one, it is an open principle and can lead to arbitrary decisions. Further, it has traditionally been applied to protect and preserve specific subjects – the Ozone layer, North Sea, biodiversity ,etc and its application to emerging tech can be tricky.⁴⁸ However, there is a spectrum when it comes to the precautionary principle, under the strong version of the principle all activities are banned unless it can be shown that there is no harm to the environment or public health. Pesticide regulation in the United States, for example, uses the strong version of the precautionary principle. Accordingly, manufacturers have to demonstrate that pesticide residue will not cause harm if they are to be allowed in food products.⁴⁹ A weak version of the precautionary principle allows the regulator to take precautionary measures based on risks to the environment but does not require it to respond. An example of the weak version of the precautionary principle is the Rio Declaration 1992, which states that, “...where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost effective measures to prevent environmental degradation.”⁵⁰ Both versions of the principle allow the state to regulate in the face of scientific uncertainty. Gregory N. Mandel and James Thuo Gathii note that there also exists **a moderate form of the precautionary principle that functions more like a norm rather**

⁴⁸ ibid

⁴⁹ Mirit Shamir et al., The application of fuzzy logic to the precautionary Principle, *Artif Intell Law* (2007) 15:411–427. “The Wingspread statement is an oft cited example of a strong precautionary principle. The statement notes that, “When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause-and-effect relationships are not fully established scientifically. In this context the proponent of an activity, rather than the public, should bear the burden of proof.” Most versions of a strong precautionary principle are not so severe. The principle effectiveness is rather because it shifts the burden of proof away from the state. Noah M. Sachs, *Rescuing the Precautionary Principle from its Critics*, *Illinois L. Rev* 1285- 1138 (2011)

⁵⁰The Ocean Foundation. (rep.). *Understanding and Applying the Precautionary Principle to Deep Sea Minerals Mining in the Pacific Islands Region: A Socio-cultural and Legal Approach*. Retrieved from <https://www.oceanfdn.org/sites/default/files/Understanding%20and%20Applying%20the%20Precautionary%20Principle%20to%20Deep%20Sea%20Minerals%20Mining%20in%20the%20Pacific%20Islands%20Region.pdf>; Marko Ahteensuu & Per Sandin, The Precautionary Principle, In: Roeser S., Hillerbrand R., Sandin P., Peterson M. (eds) *Handbook of Risk Theory*, https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-1433-5_38

than a strict legal obligation and which creates a political space for the discussion on the impacts on health, environment etc.⁵¹

Astra Taylor and Meredith Whittaker emphasize that we have to think beyond technocratic, market based solutions (such as anti-trust solutions to break up Big Tech) or individualistic solutions (for example, boycotting apps and services) when addressing issues of social justice that technologies like AI amplify. To effect a transformative change, reform to the status quo through incremental policy change and amendments to the law should go hand in hand with more radical abolitionist measures.⁵² For instance, there have been arguments and efforts to ban facial recognition altogether and not just regulate its use. Facial recognition tech is not just harmful to individual privacy it also polices vulnerable communities and has a chilling effect on society that can endanger democratic values and mechanisms. While much of the opposition to facial recognition is focussed on its use by law enforcement agencies, the slip from the widely touted safety reasons for the use of facial recognition tech (to trace missing children or securely unlock a phone) to uses antagonistic to individual and group rights is quick. And so for now, there have been demands on the moratorium in its use.⁵³ **The precautionary principle can become an useful instrument to articulate the threshold of control of AI tools whose social costs outweigh the benefits. In certain cases, such as the use of facial recognition, the precautionary principle can act as a “stop and think” mechanism allowing policy makers and the public to deliberate the effects of the technology before deciding on whether to continue its**

⁵¹ Mandel, G. N., & Gathii, J. T. (2006). Cost-Benefit Analysis Versus the Precautionary Principle: Beyond Cass Sunstein's Laws of Fear.

⁵² Demystifying Big Tech with Meredith Whittaker

<https://www.thedigradio.com/podcast/demystifying-big-tech-with-meredith-whittaker/>

⁵³ Aneja, U., & Chamuah, A. (2020). *We Need to Ban Facial Recognition Altogether, Not Just Regulate Its Use*. Tandem Research. Retrieved from <https://tandemresearch.org/publications/we-need-to-ban-facial-recognition-altogether-not-just-regulate-its-use>; Ovide, S. (2020, June 9). A Case for Banning Facial Recognition. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html>.

development or dissemination at all. In other cases it can empower the regulator to intervene if it feels, for instance, constitutionally protected rights are at stake. Because of its spectrum of application, NITI Aayog should consider how it can couch the enforcement of AI principles with the precautionary principle's jurisprudential structure. Such an approach can be designed to incentivizes private actors to disclose risks of their technology to regulators (a much needed input for the oversight body like the Council for Ethics and Technology to function) by shifting the burden of proof of no harm on to them.⁵⁴

⁵⁴ Noah M. Sachs, Rescuing the Precautionary Principle from its Critics, *Illinois L. Rev* 1285- 1138 (2011)

1.4.2 Human Rights Approach

Many within civil society have also argued for a human-rights based approach to regulating AI as opposed to purely normative principles-based approaches and broad ethical guidelines. According to Mark Latonero, the proliferation of AI principles in the past few years signals a crisis of legitimacy of governing institutions, which could have negative ramifications on a global scale. In this regard, he suggests ‘human rights law can stabilize AI governance if framed as an anchor to guide AI usage’.⁵⁵ Further, Latonero argues, in order for AI to serve any kind of greater good, the least minimum requirement would be to not harm fundamental human rights and values.⁵⁶

Critiquing the development of principle-based frameworks for AI governance, a report by human rights organization Article 19 notes that the key issue with normative approaches reliant on broad principles is that they are not actionable. AI principles often lack ‘concrete and narrowly defined language, and have failed to develop ‘independent oversight or enforcement mechanisms’.⁵⁷ In contrast, **human rights-based approach identifies “rights holders (people who use or are affected by technologies) and duty bearers (companies or governments deploying said technologies)**. It is a universal set of principles, has binding effect, and is based on the rule of law. It draws on an internationally recognised system of law that defines both business and state responsibility and the specific standards they must adhere to. This means that rights, reasonable restrictions, their status under law and implementation in practice, are anchored in a system that is verifiable, specific, and detailed.”⁵⁸

⁵⁵ Latonero, Mark (2020) *AI Principle Proliferation as a Crisis of Legitimacy*. Carr Center Discussion Paper Series. Fall 2020

⁵⁶ *ibid*

⁵⁷ Article 19. Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence, April 2019

⁵⁸ *ibid*

McGregor et. al. also argue that what makes human rights approach valuable is that it is already cognizant of the nexus that can develop between state and industry interests. The United Nations Guiding Principles on Business and Human Rights (UNGPs) recognise the possibility of states shirking their obligations under international law by contracting with private actors, and provide mechanisms to ensure the protection and fulfillment of human rights obligations.⁵⁹

However, they note, that while **states are obligated to prevent human rights violation, the same does not extend to private companies and organizations. This is problematic in the context of AI, as both the greatest amount of AI development and deployment are concentrated in the private sector. The fact that global businesses operate across multiple jurisdictions inevitably gives rise to regulatory and enforcement gaps and inconsistencies.**

Stark et. al. also highlight some of the criticisms of the approach - **particularly, the critique that the ‘minimum standards’ requirement of human rights does not account for distributive and relational justice and equality.**⁶⁰

1.4.3. Algorithmic Impact Assessments

Wide-scale adoption and deployment of algorithmic systems in recent times, have also witnessed a rise in calls for transparency and accountability with respect to AI, particularly in the case of algorithmic decision-making. In this regard, algorithmic impact assessments have become increasingly proposed as a mechanism for ensuring algorithmic accountability. For instance, the Canadian Government has developed an algorithmic impact assessment framework which is required to be used by government agencies that employ algorithms. Similarly, Algorithmic Accountability Act

⁵⁹ McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. *International & Comparative Law Quarterly*, 68(2), 309-343.

⁶⁰ Stark, L., Greene, D., & Hoffmann, A. L. (2021). Critical Perspectives on Governance Mechanisms for AI/ML Systems. In *The Cultural Life of Machine Learning* (pp. 257-280). Palgrave Macmillan, Cham.

2019 introduced in the U.S. congress also would require large companies to conduct impact assessments of automated decision systems. While several countries and organizations have proposed and are developing algorithmic impact assessments, impact assessments are not restricted to the AI sector alone. For example, aside from a long history in the environment sector, in recent times, the GDPR has also mandated privacy impact assessment.

In AINow's response to the EU white paper, it notes that rather than build a regulatory framework based on a rigid threshold between low and high risks; regulations on AI should be centred on the impact of AI instead. This is because there is high levels of discretionary decision-making in deciding what counts as a risky application. Further, AI systems can produce predictions, classifications, and decisions that may appear inconsequential at the time, but that can have lasting effects on people's lives, especially if they are used in assessments or means testing in the future. Therefore, they argue that the threshold condition for regulatory intervention in AI should be determined by the nature of the impact and the particular effects of the technology, irrespective of the sector in which they are being used.

Furthermore, **impact based assessments of AI can also help account for a wider range of risks - particularly, those that are felt at a wider societal and environmental level rather than just individuated risks of particular products and services. Even if risk-based approaches serve as an entry-point to assess and identify high risk cases,, "algorithmic impact assessments" (AIA) should be conducted to support democratic participation and accountability for the decisions of whether and how AI or algorithmic decision systems should be used.**⁶¹

⁶¹ Kak, A., Richardson, R., Dobbe, R., (June 14 2020) Submission to the European Commission on "White Paper on AI - A European Approach" AI Now Institute, New York University

However, thinking on algorithmic impact assessments is still in process of development, according to Moss et. al, there are still several open questions in relation to them. For instance, they note, what counts as an impact is non-obvious and different types of impact come into focus depending on when an assessment occurs.⁶² **While impact assessments are still being deliberated and shaped, with regard to AI, it can be beneficially applied for both ex-ante and ex-post monitoring of AI applications.** The method, as Moss et. al. note, is more attuned to the process of iterative development as well as deployment of AI - as the impacts that are discernible at the design and specification phase of a project are different than the impacts that become visible in other phases. **Further, post-deployment impact assessments may be most useful as a form of transparency for technical or historical records.**⁶³

1.4.4 Product Liability Frameworks

Product liability regimes are useful instruments to protect consumers once a product enters the market. Fault based liability regimes, as traditionally constructed however, do not translate well when applied to AI products, but scholars and policymakers are over the past few years engaged in thinking through how these frameworks can be moulded to address issues arising from tech like AI.

In India, The Consumer Protection Act, 2019 which is set to replace the 1986 Act makes major additions to the product liability regime in the country. The new Act defines product liability as “the responsibility of a product manufacturer or product seller, of any product or service, to compensate for any harm caused to a consumer by such defective product

⁶² Moss, E., Watkins, E. A., Metcalf, J., & Elish, M. C. (2020). Governing with Algorithmic Impact Assessments: Six Observations. *Available at SSRN.*

⁶³ Moss, E., Watkins, E. A., Metcalf, J., & Elish, M. C. (2020). Governing with Algorithmic Impact Assessments: Six Observations. *Available at SSRN.*

manufactured or sold or by deficiency in services relating thereto.”⁶⁴ The act lays down that an aggrieved party has cause of action against the product manufacturer, product seller or product service provider as the case maybe, if harm was suffered. Harm includes damage to property, bodily harm, and mental agony. Indian courts, however, have been reluctant to award damages for mental trauma and distress.⁶⁵ Further the plaintiff has to show that the harm was related to a defect in the product or design, inadequate instructions or warnings, lack of reasonable care, negligence, etc.⁶⁶ Apart from the Consumer Protection Act, The Motor Vehicle Act, Bureau of Indian Standards Act, Drugs Act read with Medical Devices Rules 2017 all give the regulator body the power to recall defective products.⁶⁷

The lack of transparency, explain-ability and the inherent nature of the functioning of AI systems, can make it difficult for plaintiffs to establish a causal link between defect and damage that product liability requires.⁶⁸ The AI Now Institute notes that harms from AI systems “are rarely because of the physical or technical product features alone. More often they stem from a lack of a safety culture around human-machine interactions, inadequate specifications in the engineering development process, or a lack of empirically verified safety assurances.”⁶⁹ The way autonomous systems operate means that harm is not always predictable or foreseeable.⁷⁰ Moreover, harm need not arise because of defects in the product, but as a result of the AI’s probabilistic assumptions. One way around this is to treat these systems as

⁶⁴ The Consumer Protection Act, 2019

⁶⁵ Bajaj, V., Raghavan, K., & Kaul, S. (2020, May). India - The Product Regulation and Liability Review - Edition 7 - TLR. Retrieved January 15, 2021, from <https://thelawreviews.co.uk/edition/the-product-regulation-and-liability-review-edition-7/1225197/india>

⁶⁶ Supra note 64.

⁶⁷ Supra note 62

⁶⁸ Kakkar, J & Singh, N. (August 2020) Comments to the NITI Aayog on the working document: Towards responsible AI for All. Retrieved from: <https://ccgdelhi.org/wp-content/uploads/2020/08/CCG-NLU-Comments-to-NITI-Aayog-on-the-Towards-Responsible-AI-for-All-Documents.pdf>

⁶⁹ AI Now- Submission to the European Commission on “White Paper on AI - A European Approach” <https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf>

⁷⁰ Supra note 44.

artificial persons with moral agency but as Karni A. Chagal-Feferkorn argues, autonomy is a troublesome classifier and can lead to absurd and inconsistent results.⁷¹

Another approach could be to impose strict liability for AI products that holds the producer or service provider liable even when there is no fault established. However, since multiple actors are involved in the development of AI systems, ascribing fault to one producer, service provider etc can be difficult. To overcome the ‘many hands problem’ of AI, some have suggested collective responsibility, where resources of the many actors involved in creating the AI product are pooled to compensate those harmed. Since the threshold of compensation is lowered in such a situation, there is a danger that some actors may decide to forgo taking safety and security measures. To counteract this, those actors who demonstrate a “good faith effort” to mitigate harm can be used to reduce damages paid.⁷² Further harm under the Consumer Protection Act needs to be opened up to include violation of dignitary rights and human rights.⁷³

Jurisdictions like the EU have set up bodies like the Group of Experts on Liability and New Technologies to reconcile existing liability systems with new technological systems like AI. In its report “Liability for AI and other emerging digital technologies”, the Group of Experts listed duties of “operators” of AI systems, whose infringement will induce that fault-based liability. These duties include a) choosing the right system for the right task and skills, b) monitoring the system and c) maintaining the system. Further while the Report maintains that the burden of proof should be on the

⁷¹ Karni A. Chagal-Feferkorn, Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers, 30 Stan. L. & Pol’y Rev. 61 (2019)

⁷² Stanford: Input on the European Commission White Paper “On Artificial Intelligence – A European approach to excellence and trust”, https://hai.stanford.edu/sites/default/files/2020-07/HAI_WhitePaper_v4B.pdf

⁷³ Ibid

plaintiff, it recognizes in certain circumstances where there exists disproportionate difficulties in establishing that a standard of care was breached, this burden can be reversed.

On its own product liability is ill suited to address the host of issues that AI instigates. For one, product liability kicks in only post-deployment, but importantly, it is limited to individual cases rather than the whole ecosystem of AI regulation. **Product liability is still a useful regime to address some aspects of AI regulation and should not be abandoned. We recommend that like the EU, the new oversight body, study how the Consumer Protection Act interacts with products and services that incorporate technologies like AI, and make recommendations as to how the product liability framework can be applied to emerging technologies.**

PART II: EVALUATING THE COUNCIL FOR ETHICS AND TECHNOLOGY

The NITI Aayog has proposed that the risk based regulation of AI in the country will be managed by the Council for Ethics and Technology. **The Council, however, is missing much statutory power that would give it some regulatory bite, and which traditionally, oversight bodies have been accorded. There are also key pieces of information missing from the Working Document, related to the location of the oversight body and its relationship with sectoral regulators that need to be provided if we are to assess the extent of influence and independence of such a body. Further we note that features of the modern regulatory regime of transparency, accountability, consultation and reason giving are absent from the working document and need to be expounded on.**

2.1 Understanding the Nature and Functions of Regulatory Oversight Bodies

Regulatory oversight bodies (ROBs) have become a popular intuitional arrangement in systems that have adopted the practice of Regulatory Impact Analysis⁷⁴ (RIA). These bodies are usually set up “in order to reduce the costs and ancillary risks, increase the benefits, promote cost-effective instrument choice, encourage consistency and transparency, ensure accountability and equity, and improve the overall social outcomes of Regulation.”⁷⁵ ROBs overlook the functioning of other regulators and ensure that any regulation produced has a sound economic rationale, i.e, complies with ex ante and ex-post

⁷⁴ “RIA is a process of systematically identifying and assessing the expected effects of regulatory proposals, using a consistent analytical method, such as benefit/cost analysis.” OECD. (2008). (rep.). *Introductory Handbook for Undertaking Regulatory Impact Analysis (RIA)* (1st ed.). Retrieved from <https://www.oecd.org/gov/regulatory-policy/44789472.pdf>

⁷⁵ Amo, P. A., & Rodrigo, D. (2007). *OVERSIGHT BODIES FOR REGULATORY REFORM* (Rep.). OECD. Retrieved from <https://www.oecd.org/mena/governance/38403668.pdf>

RIAs. For the ROB to be able to perform such a function, communication channels with sectoral regulators must be well established.⁷⁶ Thus, these bodies are usually (but not necessarily) centrally located and act as umbrella intuitions. Across the OECD countries ROB's have been situated in: industry, economics or commerce bodies, budgeting and auditing offices and even cabinet offices. The closer to the center the ROB is located, greater is its influence. However, the first level of RIA is still conducted by the line ministry, and so to a large degree the process is decentralized. Though, in many of the countries, guidance documents are provided by the ROB to line ministries.⁷⁷

To reach carefully deliberated decisions, ROB's need to have sufficient time. They also should strive to provide well-defined recommendations based on which regulators can take concrete steps. Further, ROB's should have the autonomy to recommend that regulators reconsider their decision and work with them in an advisory position to ensure capacity building.⁷⁸ The OECD categorises the functions of ROB's broadly as follows: quality control (scrutiny of process), identifying areas of policy where regulation can be made more effective (scrutiny of substance), systematic improvement of regulatory policy (scrutiny of system), coordination (coherence of the approach in the administration), and lastly guidance, advice and support (capacity building in the administration).⁷⁹

⁷⁶ Ibid

⁷⁷ OECD. (1997). *Regulatory Impact Analysis: Best Practices in Oecd Countries.*, <http://www.oecd.org/regreform/regulatory-policy/35258828.pdf>

⁷⁸ STAROŇOVÁ, K. (2017). Comparing the Roles of Regulatory Oversight Bodies in Central and Eastern European Countries. *European Journal of Risk Regulation*, 8(4), 723–742. <http://doi.org/10.1017/err.2017.56>

⁷⁹ OECD. (rep.). *The Role of Oversight Bodies in Promoting Good Regulatory Practices: Key Findings from the Regulatory Policy Outlook 2018.* APEC. Retrieved from http://mddb.apec.org/Documents/2018/EC/EC%20CONF/18_ec_conf_016.pdf

Due to the nature of their functions ROBs are usually staffed with technical experts, particularly but not limited to economists. It has been argued that the use of experts is also one way to maintain the independence of the oversight body⁸⁰, besides which expertise is also used to signal legitimacy.

⁸⁰ Supra note 65

2.2 Unpacking the Relationship between the Oversight Body and Sectoral Regulators

The oversight body proposed by the Working Document takes much inspiration from the institutional set up to enforce RIAs particularly instituting oversight bodies that are staffed with experts, to interface across ministries. Although India has independent regulators who were indented to work at arm's length from politics, towards technocratic solutions (something like a RIA), it never did set up an oversight body, and so the NITI Aayog's proposed institutional arrangement for the regulation of AI has no precedent it can learn from. Therefore, greater detail on the role of the oversight body in relation to guiding the risk assessment process is required.

One of the stated functions of the oversight body, i.e, Council for Ethics and Technology, will be to coordinate approaches across various regulators and assist regulators in identifying risks. Further, an ethics committee **may** be constituted to conduct the risk assessment (assess the potential for harm and benefits) and lay out a plan for risk mitigation. **Traditionally, however, it is the line ministry that makes the first level of assessment (with assistance from the oversight body) after which the assessment is evaluated by the oversight body. In some countries as indicated above, oversight bodies have the power to reject a risk assessment conducted by a line ministry. The advantage of the latter model is that it accounts for the independence of the sectoral regulator.** Besides which, since the NITI Aayog Working Document presumably relies on the enforcement powers of sectoral regulators⁸¹, allowing the latter to conduct the first level of assessment could enrich capacity building, especially since many of the sectoral regulators in India do not have the experience

⁸¹ The Working Document notes that "Existing regulatory instruments are best placed to enforce rules, standards and guidelines"

of governing newer technologies like AI. Additionally, since the oversight body also overlooks private actors, guiding risk assessments for both sectors may be much too burdensome for the body.⁸²

In other jurisdictions an institutional ecosystem has been created that manages the many facets of regulating digital technologies including AI. The UK has an Office of Artificial intelligence (which is part of the Department for Digital, Culture, Media & Sport and the Department for Business, Energy & Industrial Strategy), that issues relevant guidelines (for example, it has issued Guidelines for AI Procurement⁸³ by public sector bodies and a Guide to Using Artificial Intelligence in the Public Sector,⁸⁴ amongst others.)⁸⁵ The UK also has a Centre for Data Ethics and Innovation (CDEI) which is an independent advisory body (Under Department for Digital, Culture, Media & Sport) tasked with analysing the opportunities and risks of data driven technologies.⁸⁶ Spain has a Secretary of State for Digitalization and Artificial Intelligence which operates under the Ministry of Economic Affairs and Digital Transformation.⁸⁷ The country also recently constituted an Artificial Intelligence Advisory Council whose key members include the Minister of Economic Affairs and Digital Transformation as well as the Secretary of

⁸² Although some of this burden may be alleviated by the ethics council, the Working Document notes that they ‘**may**’ review procurement of AI by public authorities.

⁸³ *Guidelines for AI procurement*. GOV.UK. <https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement>.

⁸⁴ Office for Artificial Intelligence. (rep.). *A guide to using artificial intelligence in the public sector*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871178/A_guide_to_using_AI_in_the_public_sector_web_version.pdf

⁸⁵ GOV.UK. (2020, March 25). *Guidance and regulation*. GOV.UK. <https://www.gov.uk/search/guidance-and-regulation?organisations%5B%5D=office-for-artificial-intelligence&parent=office-for-artificial-intelligence>.

⁸⁶ *Centre for Data Ethics and Innovation*. GOV.UK. <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>.

⁸⁷ *Spanish Secretary of State for Digitalization and Artificial Intelligence (Spain)*. ZSI. Retrieved from <https://www.zsi.at/en/object/partner/5746>.

State for Digitalization and Artificial Intelligence.⁸⁸ We could also benefit from the above approaches, which seeks to set up a constellation of state actors that manages various aspects of regulating AI.

India is only now setting out to contend with the regulatory issues thrown up by emerging tech, consequently, the regulatory landscape is still marked by critical lacunas. **In such a case leaving the oversight body without any bite is a missed opportunity for concrete regulation.** Other civil society organizations have also echoed such a proposal, the Internet Freedom Foundation recommends that the Council for Ethics and Technology must have regulatory powers akin to that which is envisioned for the Data Protection Authority under the Personal Data Protection Bill, 2019⁸⁹ Regulatory powers are also important if the Council for Ethics and Technology is to satisfactorily carry out its duties, for instance it will need to request/demand information from both public and private actors, and will need regulatory powers to do so.⁹⁰ Similar charges were levied against the CEDI which also only performs an advisory function.. Scholars have argued that the need for a cross-sectoral regulator for emerging technology like AI will arise, and it would be wise to equip the oversight body when it is established rather than wait to do it a later time.⁹¹ In the case of the UK however, the CEDI is envisioned to become a statutory body, in which case it will eventually also have statutory powers.⁹²

⁸⁸*El Gobierno constituye el Consejo Asesor de Inteligencia Artificial - Ministerio de Asuntos Económicos y Transformación Digital (es)*. Vicepresidencia Tercera del Gobierno - Ministerio de Asuntos Económicos y Transformación Digital. Retrieved from <https://www.mineco.gob.es/portal/site/mineco/menuitem.ac30f9268750bd56a0b0240e026041a0/?vgnnextoid=51884ba89bc63710VgnVCM1000001d04140aRCRD>.

⁸⁹ Garg, R. (2020, December 14). *Going beyond hashtags: how to ensure AI technology truly benefits everyone*. Internet Freedom Foundation. <https://internetfreedom.in/ensuring-true-ai-for-all/>.

⁹⁰ The Alan Turing Institute. (rep.). *Response to the DCMS consultation on the Centre for Data Ethics and Innovation*. Retrieved from https://www.turing.ac.uk/sites/default/files/2018-09/response_of_the_alan_turing_institute_to_dcms_cdei_consultation.pdf

⁹¹ Whittlestone, J., & Stix, C. (2018). (rep.). *Response to the Consultation on the Centre for Data Ethics and Innovation*. Retrieved from http://lcfi.ac.uk/media/uploads/files/Whittlestone_Stix_CDEI_consultation_response_-_June_2018_copy_LtYK2IJ.pdf.

⁹²HAT Community Foundation. (2018). (rep.). *Response to Centre for Data Ethics and Innovation Consultation*. Retrieved from <http://wrap.warwick.ac.uk/114600/>

In countries that do not want only an advisory oversight body, efforts have been made to strengthen the existing regulatory landscape and buffer it through the addition of new regulators when required. The German Data Ethics Commission advises a distributed supervisory role amongst sectoral regulators but has recommended that in order to prevent any gaps in supervision, the government identify areas where there are no sector-specific expertise available. Once identified it notes that the state may find that existing regulators can be assigned new areas where compatible (especially the data protection authority). However, it notes there may still be a need to establish new regulatory bodies when required.⁹³ On a similar note, the Danish Government intends to set up an inter-ministerial working group to identify gaps in the existing legislative framework and determine the need for new legislative initiatives to address AI governance. India still does not have a Data Protection Authority, a body most suited to shoulder some of the regulatory burden with regard to emerging tech, and therefore it is critical that we have a strong oversight body in this field. **We thus support the effort of the oversight body in monitoring existing policies and regulations gaps and inconsistencies based on which it will make recommendations, but also suggest that some thought be given to strengthening the oversight body to provide it with not just advisory but also enforcement powers.**

2.3 Instituting Measures for Independence, Transparency and Accountability

Transparency, openness, consultation and reason-giving are characteristics of modern regulatory systems.⁹⁴

Important features of a regulatory system that serve these functions are missing from the NITI Aayog Working

⁹³ Daten Ethik Kommission . (rep.). *Opinion of the Data Ethics Commission*. Retrieved from https://www.bmju.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf?__blob=publicationFile&v=3.

⁹⁴ Thiruvengadam, A. K. (2017) Flag-bearers of a new era? The evolution of new regulatory institutions in India (1991–2016). In Rose-Ackerman, S., Lindseth, P. L., Emerson, B. (Eds.), In *Comparative administrative law* (2nd ed., pp. 218–233).Edward Elgar Publishing.

Document. Apart from which, the Working Document gives little indication of where the oversight body will be located. Would it be centrally located within the NITI Ayog, or for instance the Prime Minister’s Office (PMO)? Or would it be spread across ministries that are most impacted by AI such as Labor and Employment, Law and Justice, Minority Affairs, Transport, Social Justice and Employment, Information and Broadcasting, Commerce, and Industry, and Finance. The decision of where to situate the oversight body must be made carefully, for although situating it closer to the center, for instance the PMO would give it greater power, it would also make it easier to usurp control from the body. **Without the location of the body’s operations, its source of funding for the functioning of the body, an important component of its independence, is unknown.** In the case of the UK, the Framework Agreement between the Department for Digital, Culture, Media & Sport and the CEDI, states that the latter will be funded by the Department. The Chair of the CEDI and the Secretary of State for Digital, Culture, Media and Sport also work closely to determine CEDI’s work program and an Independent Board set up by the Secretary of State is tasked with checking the CEDI’s delivery against its objectives.⁹⁵

The Council for Ethics and Technology is expected to be staffed with experts across fields (including computer science and AI, law, sectoral experts, humanities, and representatives from industry, standard setting bodies) and civil society. **However, expertise cannot be the sole bulwark of independence. There is unfortunately no mention of the appointment procedure, tenure or remuneration, all important considerations for independence of the chairperson and members of the Council for Ethics and Technology and ethics committees.** In the Framework Agreement mentioned above, appointment procedures to the CEDI are clearly laid out and so are the interactions between CEDI and government bodies.⁹⁶

⁹⁵Department for Digital, Culture, Media & Sport, Centre for Data Ethics and Innovation. (2019). (rep.). *Framework Agreement between the Department for Digital, Culture, Media & Sport and the Centre for Data Ethics and Innovation* . Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/788739/Framework_Agreement_with_CDEI.pdf

⁹⁶Ibid

Further, as important it is for the oversight body to have defined relationships with regulatory agencies, it is also necessary to protect the body from being used instrumentally by regulated entities (private actors) to influence sectoral regulators.⁹⁷

On the question of openness and transparency, again provisions have been made in the Framework Document expressly subjecting CEDI to The Freedom of Information Act 2000, likewise the proposed Council for Ethics and Technology and ethics committees that may be constituted to review public procurement must be subject to the Right to Information Act, 2008. In a step towards transparency, the Working Document states that every review must be documented, including the risks identified, mitigation strategy, and comments from the committee members. Whether this will be accessible by the general public, however, has not been stated. In the UK, apart from publishing its annual work program and report, the CEDI also makes public recommendations and actions taken by the government related to data and AI, particularly action taken to address barriers to ethical use. If recommendations cannot be made public in the UK, the Framework Agreement states that the reason must comport with a legislative exception to access to information.⁹⁸

Public participation is an important component of democratic decision making and can go a long way in engendering trust. The Working Document does specify engaging with local communities and civil society to monitor the impact of the deployment of AI on them. There is unfortunately little detail about how local communities will be engaged.

Emanuel Moss et al note that badly designed public engagement strategies can “flatten asymmetries in agency, power, voice,

⁹⁷ Cohen, J. E. (2016). The Regulatory State in the Information Age. *Theoretical Inquiries in Law*, 12(2), 369–414. <https://doi.org/10.1515/til-2016-0015>

⁹⁸ Supra note 94

and vulnerability.” If public consultation is not meaningfully designed at best it is a farcical performance of care, but at worst can end up endangering vulnerable populations by making their vulnerability legible to bad actors.⁹⁹

2.4 Unpacking the use of standards and self -regulation by the private actors

To regulate the private sector, the NITI Aayog Working Document suggests using a combination of self-regulation and mandatory compliance in cases of ‘high-risk’. **Leaving aside the complexities of defining what is ‘high-risk’, the reliance on self-regulation, through the adoption of standards, can easily slip into no-regulation.** Professor Julie Cohen observes that the regulatory state in the information age is typified by technical standard setting, unfortunately the latter are often secretive and averse to public scrutiny.¹⁰⁰ **The overwhelming focus on standards set by technical bodies also does not hold up well against rule of law traditions which require the law to be publicly made, reasoned, contestable, and transparent to those it affects.** The state here is displaced from its legislative capacity and reduced to a managerial role to steward the processes of privatized regulation.¹⁰¹ Compliance with standards, Cohen notes can be easy to establish, but what that means substantively can be hard to translate to the public or even regulators.¹⁰² We urge that the oversight body evaluate the processes of standard-setting bodies for public engagement and inclusivity prior to adopting any standards developed by them. Technical systems tend to make abstract values such as fairness and justice to such an extent that they lose meaningful

⁹⁹ Moss, E., Metcalf, J., & Elish, M. C. Governing with Algorithmic Impact Assessments: Six Observations. *SSRN.com*.
<https://doi.org/https://dx.doi.org/10.2139/ssrn.3584818>

¹⁰⁰Supra note 97.

¹⁰¹Berkman Klein Center. (2019). (rep.). *Between Truth and Power: The Legal Constructions of Informational Capitalism* . Retrieved from
https://cyber.harvard.edu/sites/default/files/2019-12/2019_11_21_Berkman_Julie_Cohen_NS.pdf

¹⁰²Supra note 97.

application.¹⁰³ The design of AI systems to account for fairness, accountability and transparency may have once been a technical endeavour, but increasingly lawyers, philosophers and social scientists, amongst others, are weighing in to push technical systems to account for socio-political dimensions,¹⁰⁴ or to argue against technical solutions in the first place.¹⁰⁵

Self-regulatory mechanisms which may operationalize these standards also suffer from a democratic deficit. These mechanisms often lack transparency, accountability and consultation with interested parties that are not part of industry. Many also consider self-regulation, such as Facebook’s Oversight Board, an attempt to “stave off actual government regulation” that is designed to serve narrow business interests or worse just a public relations mechanism to fend off actual responsibility to users.¹⁰⁶

If self-regulation is to be seen as a legitimate process, it must include some degree of public participation and a commitment to enforcement.¹⁰⁷ Further, self-governance mechanisms should be subject to the purview of an independent body and should set up mechanisms for internal and external transparency. Gijs Leenders, for instance, notes that the

¹⁰³ Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2018). ACM Conference on Fairness, Accountability, and Transparency . https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265913.

¹⁰⁴ Article 19. (2019). (rep.). *Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence*. Retrieved from https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf

¹⁰⁵ Supra note at 103

¹⁰⁶ Klonick, K. (2020). The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, 129(8).

¹⁰⁷ Price, M., & Verhulst, S. The concept of self-regulation and the internet. In J. Waltermann & M. Machill (Eds.), *Protecting our children on the internet: Towards a new culture of responsibility* (pp. 133–198). essay, Bertelsmann Foundation Publishers.

Partnership on AI (PAI) has no mechanisms in place to monitor compliance with its principles or means of enforcement. This is to be expected since many of the members of PAI are industry rivals.¹⁰⁸

In conclusion, we resonate with Professor Cohen conclusions that while it is necessary that regulatory mechanisms evolve to keep pace with emerging technology, **issues related to the lack of accountability and effective governance are related to institutional design, and the public must have a say in such design.**¹⁰⁹ While we are aware of the problem of regulatory burden, especially on start-ups and SMEs, we recommend that NITI Aayog flesh out mechanisms to buttress private sector regulation through some oversight by the state (independent) across risk profiles. For example, Denmark has a legislative proposal that mandates private companies to disclose their data ethics policy and even institutes fines for non-compliance.¹¹⁰

Not all ethics oversight bodies work on the evaluation of risk. The terms of reference for CEDI states that the Centre will: 1) Analyse and anticipate risks and opportunities; 2) Agree and articulate best practice; and 3) Advise on the need for action.¹¹¹ Though the CEDI is expected to take decisions that are based on evidence and account for economic impacts of a regulation it is not constrained to a risk-based analysis.¹¹² Singapore's Advisory Council on the Ethical Use of Artificial Intelligence with its dedication to private sector participation towards innovation and consumer confidence is also not committed to a risk based

¹⁰⁸ Leenders, G. (2019, April 23). *The Regulation of Artificial Intelligence-A Case Study of the Partnership on AI*. Medium. Retrieved from <https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f>.

¹⁰⁹Supra note 105.

¹¹⁰ *Denmark introduces mandatory legislation for AI and Data Ethics*. 2021.AI. (2020, October 15). <https://2021.ai/denmark-introduces-mandatory-legislation-ai-data-ethics/>.

¹¹¹ Centre for Data Ethics and Innovation. (2019, July 2). *Centre for Data Ethics and Innovation (CDEI) Terms of Reference*. GOV.UK. <https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-cdei-terms-of-reference/centre-for-data-ethics-and-innovation-cdei-terms-of-reference>.

¹¹² Ibid

system to develop ethics standards, publish advisory guidelines, practical guidance, and/or codes of practice for the voluntary adoption by the industry.¹¹³ Furthermore, even when the ethics oversight body does adopt a risk based approach – it is not merely or even primarily an economic calculus designed to ensure that innovation is not deterred, but also an effort to protect legal rights including human rights and democratic principles. The Germany’s Data Ethics Commission recommends a risk-based approach to AI regulation, but uses a more holistic assessment of risk that accounts for ecological, social, psychological, cultural, economic and legal dimensions.¹¹⁴ Though the NITI Aayog working document adopts a risk based approach, we urge that it expands the notion to risk beyond efficiency metrics and accounts centrally for qualitative considerations of redistributive justice, human rights and constitutional principles. While a regulatory system may be tasked with conducting stringent cost-benefit analysis prior to initiating regulatory action, Dubash and Morgan reflect that in reality these regulators often have to make room for political considerations that are often legitimate. The role of the regulator then is not to shield decision-making from politics but to create synergies between the regulatory state and the political environment. “In other words, the regulatory state should play a role in creating the conditions for political rather than only technocratic and apolitical deliberation.” Such an approach also allows regulators to factor in considerations for redistribution, that are particularly salient in the global South¹¹⁵

¹¹³ *Inaugural meeting of the Advisory Council on the Ethical Use Of Artificial Intelligence and Data*. Infocomm Media Development Authority. [https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/inaugural-meeting-of-the-advisory-council-on-the-ethical-use-of-artificial-intelligence-and-data](https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/inaugural-meeting-of-the-advisory-council-on-the-ethical-use-of-artificial-intelligence-and-data;); Infocomm media development authority. (2018). (rep.). Artificial Intelligence Governance and Ethics Initiatives. Retrieved from Inaugural meeting of the Advisory Council on the Ethical Use Of Artificial Intelligence and Data. Infocomm Media Development Authority. <https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/inaugural-meeting-of-the-advisory-council-on-the-ethical-use-of-artificial-intelligence-and-data>.

¹¹⁴ Supra note 93.

¹¹⁵ Dubash, N. K., & Morgan, B. (2013). The Embedded Regulatory State: Between Rules and Deals. In N. K. Dubash & B. Morgan (Eds.), *The Rise of the Regulatory State of the South Infrastructure and Development in Emerging Economies* (pp. 163–184), Oxford University Press.

PART III: MOVING TOWARDS ANTICIPATORY GOVERNANCE AND RESPONSIBLE RESEARCH AND INNOVATION

3.1 Introduction to Responsible Research and Innovation

The scientific and research community has also been supporting arguments in favor of shifting from the “governance of risk” to the “governance of innovation” through “a more collective stewardship of science and innovation”¹¹⁶

This idea of anticipatory governance emerged as a reorientation towards upstream governance in the early 2000s. Anticipatory governance implies that effective action is based on more than just sound analytical capacities and relevant empirical knowledge. It emerges out of a distributed collection of social and epistemological capacities, including collective self-criticism, imagination, and the disposition to learn from trial and error. Anticipation implies an awareness of the co-production of socio-technical knowledge and the importance of richly imagined socio-technical alternatives that might inspire its use.¹¹⁷ Versions of upstream governance such as the Responsible Research and Innovation (RRI) have picked up steam, as a consequence of waning public trust in science and innovation.¹¹⁸ Here responsibility functions as a value that guides institutional design rather than a rule.

¹¹⁶ Stilgoe, J., Owen, R., & Macnaghten, P. (2020). Developing a framework for responsible innovation. *The Ethics of Nanotechnology, Geoengineering and Clean Energy*, 347–359. <https://doi.org/10.4324/9781003075028-22>.

¹¹⁷ Burget, M., Bardone, E. & Pedaste, M. Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review. *Sci Eng Ethics* 23, 1–19 (2017). <https://doi.org/10.1007/s11948-016-9782-1>

¹¹⁸Thapa, R. K., Iakovleva, T., & Foss, L. (2019). Responsible research and innovation: a systematic review of the literature and its applications to regional studies. *European Planning Studies*, 27(12), 2470–2490. <https://doi.org/10.1080/09654313.2019.1625871>

Innovation when applied at scale necessarily changes settled ways of doing things. **Evelien de Hoop et al explain that “innovation can only be responsible if it takes into account what has to be sacrificed during innovation”.** ¹¹⁹

3.2 Components of RRI

Early articulations of RRI recognized that there was a need to evolve from “science in society to ‘science for society, with society’. It argued that equal attention should be paid to the purposes and intentions behind innovation and what we want science and innovation to do, as is paid to understanding the benefits and risks of the intended and unintended consequences of the products of innovation. **RRI notes that deciding the purpose of innovation is a deeply political question, and advocates for democratic deliberations around it. These processes are further to be institutionalized in science and innovation policies so that innovation can be guided to socially desirable ends.**¹²⁰ Richard Owen et al., identify three main characteristics of RRI: **anticipation** - to investigate socio-economic, environmental and other impacts; **reflectivity** - to gain an awareness of underlying purposes, assumptions made, risks assumed etc; and **deliberation** - on the process of reflection through engagement with the public and relevant stakeholders. The authors also note that RRI “confers new responsibilities not only on scientists, universities, innovators and businesses, but also on policy makers and research funders.” As characterized by Owen, responsibility is not a consequentialist rule but a value to be realized in “in the context of innovation as a future-oriented, deeply uncertain, often complex and always collective phenomenon.” ¹²¹ **RRI envisions a role**

¹¹⁹ de Hoop, E., Pols, A., & Romijn, H. (2016). Limits to responsible innovation. *Journal of Responsible Innovation* , 3(2), 110–135.

<https://doi.org/https://doi.org/10.1080/23299460.2016.1231396>

¹²⁰Owen, R., & Macnaghten, P. (2012). Responsible Research and Innovation: From Science in Society to Science for Society, with Society. *Science and Public Policy*, 39(6), 751–760. <https://doi.org/10.1093/scipol/scs093>

¹²¹Ibid

for public and civil society participation throughout the innovation cycle and elevates social, ethical and environmental impacts to that of commercial and technical factors in assessing innovation.¹²²

The governance and management of risk to contain negative consequences has been the predominant approach of tech policy, but RRI seeks to align itself with what has traditionally been public policy's focus i.e. searching for positive rights impact. René von Schomberg notes that the key question then becomes: what would be the "right" impacts of research and innovation policy. **RRI challenges the assumption that investing in research and innovation will automatically benefit society, instead innovation needs to be actively steered towards meeting socially beneficial ends.** Schomberg then proposes two dimensions to RRI- product and process. With regard to the product dimension (in the context of the EU) the following are identified as lodestones of innovation: fundamental values of the EU charter on fundamental rights, EU's objective of sustainable development, and the normative values embedded in "the Treaty on the European Union, such as "quality of life", "equality among men and women" etc." ¹²³

The process dimension advocates for a participative approach to innovation through which innovators become responsive to the public.¹²⁴ Various methods of public participation such as, consensus conferences, public hearings and citizen forums in research and innovation found a place in the 2001 White Paper from the EU on "European Governance". More recently,

¹²² Landeweerd, L., Townend, D., Mesman, J., & Hoyweghen, I. V. (2015). Reflections on different governance styles in regulating science: a contribution to 'Responsible Research and Innovation.' *Life Sciences, Society and Policy*, 11(1). Retrieved from <https://doi.org/10.1186/s40504-015-0026-y>.

¹²³ Schomberg, R. V., & Romijn Responsible Innovation, H. (2014). The Quest for the 'Right' Impacts of Science and Technology: A Framework for Responsible Research and Innovation. In J. van den Hoven, N. Dooran, T. Swierstra, & B. Koops (Eds.), *Responsible Innovation* (Vol. 1, pp. 33–50), Springer Science.

¹²⁴ Ibid

Horizon 2020 a research and innovation funding program from the EU has incorporated the RRI framework whose major component is public engagement.¹²⁵

Promoting product and process dimensions of the RRI framework can form a part of the new Council for Ethics and Technology functions, by interfacing with government and private funders of science research.

3.2.1 Product Dimension in the Indian context

The responsible AI principles identified by the NITI Aayog, and constitutionally enshrined values promoting social and economic justice, and civil-political rights should guide research and innovation. The innovation ecosystem in the country should not only refrain from violating these principles (mitigating harm) but seek to promote them.¹²⁶

3.2.2 Process Dimension in the Indian context

Although India has witnessed popular counter-movements by citizens to the emergence of new technologies (think protests against the introduction of Bt brinjal), Yandong Zhao et al., observe that the **practice of public engagement in science and technology in the country remains underdeveloped**.¹²⁷ The Science, Technology and Innovation Policy, 2013 developed by

¹²⁵ Ibid

¹²⁶ <https://cis-india.org/internet-governance/comments-to-aiforall-pdf>

¹²⁷ Zhao, Y., Fautz, C., Hennen, L., Srinivas, K. R., Li, Q., Zhao, S., & Stemerding, Y. (2015). Public engagement in the governance of science and technology. In M. Ladikas & M. Chaturvedi (Eds.), *Science and Technology Governance and Ethics* (pp. 39–51). Springer.

the Department of Science and Technology, emphasized a shift from an input driven model to a development-led model of innovation, that included open dialogue through which ‘People and decision makers must be made aware of the implications of emerging technologies, including the ethical, social and economic dimensions.’¹²⁸ The Department is now in the process of drafting the new Science Technology and Innovation Policy, 2020.¹²⁹ The hope is that experiences of using technology to address issues arising from the pandemic will serve as a teaching point to shine light on the importance of public participation in research and innovation. We appreciate the efforts taken in the Working document to incorporate public engagement (for example, working with local communities to understand impact of AI on the ground, including members from stakeholder community (user community) in the ethics body as well as staffing the oversight body with members of civil society), and think the same should be followed in ensuring various forms of public engagement in the upstream governance of research and innovation. Ultimately if the RRI process is infructuous it could be indicative that the innovation process should be stalled.¹³⁰

¹²⁸ Ministry of Science and Technology. (2013). (rep.). *Science, Technology and innovation Policy 2013*. Retrieved from https://unctad.org/system/files/non-official-document/CSTD_2013_STI_India.pdf

¹²⁹*Science, Technology and Innovation Policy 2020*. Science Policy Forum. <http://thesciencepolicyforum.org/initiatives/science-technology-and-innovation-policy-stip-2020/>.

¹³⁰Supra note 116.



About Tandem Research

Tandem Research is an independent research collective that generates policy insights at the interface of technology, society, and sustainability. We believe that evidence based policy making along with broad based public engagement must steer India's technology transitions. Our work seeks to ensure that India's technology transitions are just and inclusive. Tandem Research is one of the first think tanks in the country to have a dedicated AI & Society program and is widely recognised as one of the prominent voices on AI ethics and governance in India.

For further information about Tandem Research, or this response, please contact Tandem Research Director, Urvashi Aneja - urvashianeja@tandemresearch.org