Developing an AI Governance Framework for the Philippines: a Report of Preliminary Stakeholder Consultations and Review of the Literature[1]

Peter A. Sy
University of the Philippines

Executive Summary

Artificial Intelligence (AI) refers to any system simulating human traits like reasoning, problem-solving, or decision-making. An AI governance framework outlines the fundamental ethical principles to guide the development of AI in the country and the practices around the use of AI technologies. Such a framework can guide organizations and individual practitioners in navigating complex ethical and governance issues. This Report is informed by the enabling mechanisms for the retooling of institutions and human resources, or lowering the barriers to the development of AI.

The Report discusses the principles for AI systems and practices, including those adopted by 42 countries (namely: inclusive growth, sustainable development and well-being, human-centered values and fairness, transparency and explainability, robustness, security and safety, and accountability). Trust is added to fill a perceived gap in those principles, especially in relation to Philippine society.

In consultation with stakeholders and the literature, the Report identifies the elements of AI governance that organizations should consider: engaged board oversight,

---

[1] Shortcut to a live version of the document: ai-ph.org/devframe. Updated: 4 Sept 2022. Send comments and suggestions to psy@up.edu.ph. Acknowledgements: Special thanks to Yousef Concepcion, Althea Marcelo, and Sheila Ruby Dellosa (University of the Philippines) as well as Ysa Sofia Robin (De La Salle University) for their assistance in preparing this Report; Rhoanna Marie Dizon for the illustrations used in this Report; Dr. Jodie Lobana (McMaster University) for her overall guidance; Usec. Rafaelita M. Aldaba, Undersecretary for Competitiveness and Innovation, Department of Trade and Industry (DTI) for initiating the project; DTI staff for ensuring we get the widest range of stakeholder participation, despite time constraints. I am most grateful to organizations, government agencies, and businesses that provided inputs to the Report. Their representations are acknowledged in Appendix C.

performance management, regulatory compliance, strategic planning, and robust governance of data assets. Good leadership and management are essential building blocks as well for the success of AI.

For AI to be successfully adopted by society, it has to be human-centric. Even for seemingly bewildering AI algorithms using (seemingly) humanly intractable number of data points to arrive at decisions and actions affecting humans, *human accountability* is essential. The roles and responsibilities of humans working on AI projects and programs need to be defined and AI regulations and ethical standards putting human accountability front and center need to be put in place. AI must not be used for manipulation and exploitation of humans. The Report supports a ban on lethal autonomous weapons systems and killer robots. Indiscriminate surveillance and AI-based social scoring of natural persons should also be prohibited.

Data management makes sense only if there's sufficient data to manage, to begin with. Hence, the Report proposes the release of massive amounts of important, timely data (especially government data) for the research and development of AI systems. The need to establish internal and external structures for organizations in order to manage shared data risks, exercise oversight, and facilitate training requirements surrounding AI development, cannot be overemphasized. There are also common risks associated with AI that need to be addressed. Such risks include (but not limited to) bias leading to discrimination, security and privacy breach, cyberattacks, and unequal access to services, platforms, and information. To deal with these risks, risk-based AI governance and regulation are needed.

In the end, AI governance and development are also driven by the political, economic, societal, technological, and legal environments. Continued studies on these areas are as important as the establishment of robust AI governance structures.

To address these issues and concerns, the Report has formulated recommendations for areas defining AI governance and development in the country, including digitization and infrastructure, workforce development, regulation, and research and development (R&D).

1. Introduction

Artificial Intelligence (AI) is an emerging technology capable of developing new industries and uplifting the Philippine society. Numerous applications across sectors (health, finance, education, and manufacturing, among others) now use AI. Whether the country is able to maximize its benefits depends on the ability of its champions to wield

AI efficiently. Doing so requires a governance framework in the deployment of AI in Philippine society (Castro, 2022).

Informed by the Philippine AI Roadmap (Department of Trade and Industry, n.d.), this Report outlines fundamental ethical principles guiding Philippine AI systems and AI-augmented practices. It identifies the important elements of a would-be Philippine AI governance framework. The Report is the result of an iterative literature review process and stakeholder consultations. “Best practices” in the development of similar frameworks in other countries are examined as references in providing a good foundation for the governance and regulation of AI in the country. Special attention is given to inhibitors and drivers of AI development in society, as no AI governance can be oblivious to such factors.

Each section (or subsection) in this Report is potentially expandable to a separate report. But efforts are made to keep “everything” in one place and temper the desire to be comprehensive with the need to convey effectively to stakeholders the essential points of developing a national AI Governance framework. Their voices and participation are also being incorporated into this report's writing.

What is AI? The Artificial Intelligence (AI) field is dedicated to developing systems capable of performing tasks and solving problems commonly associated with human intelligence (Alpaydin, 2021). It encompasses all computer systems that demonstrate the capacity to learn, adapt, and operate in dynamic or uncertain environments (Miailhe, 2018). Also considered AI is the simulation of human traits such as knowledge, reasoning, problem solving, perception, learning and planning, producing an output or decision, including prediction, diagnosis, recommendation, or classification (Singapore Governance Framework, 2020). Most (if not all) systems that make decisions normally requiring human expertise fall within the purview of AI (Shubhendu & Vijay, 2013). In AI, such systems may include the following non-mutually exclusive capabilities (Russell & Norvig, 2020):

(a) *Natural language processing* to communicate successfully with humans;
(b) *Knowledge representation* to store what it knows;
(c) *Automated reasoning* to answer questions or draw new conclusions;
(d) *Machine learning* to adapt to new conditions or to detect and extrapolate patterns.

Moreover, other researchers consider physical stimulation as a necessary demonstration of human intelligence, hence, including other capabilities such as:

(e) *computer vision and speech recognition* to perceive the world;
(f) *robotics* to move and manipulate objects.

These six capabilities are exhibited by most AI models currently being used in many countries. They tend to define the general technical areas AI pursuits.

Why framework? A governance framework enables mechanisms for organizations, solution providers, and stakeholders to have a clear understanding of expectations, objectives, risk appetite, accountabilities, and responsibilities (Smith & Brooks, 2013). Informed by shared principles and values, frameworks enable organizations to integrate a range of plans, policy choices, approaches, strategies, and procedures into a coherent guidance document. Additionally, a governance framework of rules and practices provides an essential supporting structure for technological development. That the governance of artificial intelligence (AI) would require a good framework is practically inescapable.

Germany (2018), the US (2017), the UK (2021), and Singapore (2020), among others, have laid out their AI strategies and frameworks. Germany aims for the responsible development and use of AI, serving "the good of society", and "integrating AI in society in ethical, legal, cultural and institutional terms in the context of a broad societal dialogue and active political measures" (German Federal Ministry for Economic Affairs and Energy, 2018). Of particular concern for the Americans are robo-advisers for investors and the financial services industry providing clients with investment advisory services on various digital platforms (US Securities and Exchange Commission, 2017). The UK's national AI strategy is predicated on the country's ability to attract talents from all over the world to advance AI research and innovation (UK Office for Artificial Intelligence, 2021). Singapore's AI governance framework aims "to provide detailed and readily implementable guidance to private organizations on how to address ethical and governance issues when using AI solutions" (Personal Data Protection Commission Singapore, 2020).

As with Germany and Singapore, "ethics first" guidance on governance and policies has been consistent with many other governance frameworks. For the European Commission's High-Level Expert Group on AI (2019), trustworthy AI should be *ethical*, "respecting ethical principles and values," in addition to being lawful and robust. Principle-based frameworks lay out the values that will inform a diverse range of AI systems and AI-augmented practices.

What for? In setting its national AI Roadmap, the Philippines signals the willingness (if not readiness) "to harness AI's potential to uplift Filipinos, our local industries, and our economy" (Department of Trade and Industry, n.d.). Moreover, the development of an AI governance framework contributes to the country's attempt to become a more innovative, inclusive, and resilient economy (Castro, 2022).

Deployment of AI technologies in both public and private sectors is expected to spur economic growth, hopefully improving the lives of the Filipino people (Amazon Web

Services, 2021). Comparatively, Pillar 2 of the UK National AI Strategy is about AI benefitting all sectors and regions. Similar language of AI's promise of beneficence, improvement in a people's quality of life, and economic prosperity can be found in practically all the other national AI frameworks.

In developing the Philippine framework for AI governance and determining its scope and limits, its developers and stakeholders must be cognizant of the intended utility and inevitable economic, political, and socio-cultural impact of AI in Philippine society. Jumping out of initial discussions with stakeholders is the concern that such a framework might be used to regulate AI systems and practices *prematurely*. However, questions of government regulations in many parts of the world are no longer about "if" or "when" but "how." While the EU and the UK have signaled the intent to regulate AI soon, *governance is much broader than government regulation* or government itself.

As Singapore has shown, for instance, an AI governance framework can serve as a guide for private organizations in navigating complex ethical and governance issues (Personal Data Protection Commission Singapore, 2020). It can also be used to create or strengthen enabling mechanisms, retool institutions, or remove barriers to the development of AI. The framework could be an impetus for government agencies, industries, and academia to explore or adapt to new economic realities and emergent technologies spun by AI (Castro, 2022). With an AI governance framework, organizations could be equipped with ethical and legal tools that enable increased productivity and enhanced competitiveness—which may lead to economic growth and better quality of life (Personal Data Protection Commission Singapore, 2020). Finally, an AI governance framework signals the need for parallel development of a code of conduct by AI practitioners for a progressive industry regulation.

To imbibe trust, safety, and security in the deployment of AI in Philippine society, an AI governance framework should have specified its guiding principles.

2. Principles for AI Systems and Practices

The Organisation for Economic Co-operation and Development (OECD, 2019) offers a set of value-based principles to guide the development and utilization of AI. These principles adopted by 42 countries by far include inclusive growth, sustainable development, and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; and accountability. In addition, we maintain that trust is a fundamental value and principle essential to the overall development of AI and emergent technologies.

2.1. Inclusive growth, sustainable development, and well-being. With the recent enactment of the Philippine Innovation Act (RA 11293) (Philippine Innovation Act, 2018)

that aims to fund and support research and development (R&D) efforts in the country, AI research and development cannot be divorced from sustainable and inclusive growth. Our national developmental goals have to be intertwined with the competitiveness of micro, small and medium Enterprises that in turn have to advance the "inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and [to protect] natural environments" (OECD, 2019). AI's burdens and benefits must be shared equitably, always seeking to empower and engage everyone. National strategies for AI development should always take into account both the advantages and the risks associated with the deployment of AI. Mitigating its potential negative effects on society is essential (Amazon Web Services, 2021).

There is tacit, wide acceptance of the principle of inclusive growth, sustainable development and well-being, if the consultations with AI startups, the academe, civil society groups, big technology companies, and the government were any indication. To begin with, there is a lack of national industries that arguably tend to impede AI growth (see Section 8 and Appendix G). The IBON Foundation (2016) maintains that the lack of such industries has created structures undermining inclusive economic growth of the Philippines. Stakeholders are also well aware of widespread dominance by foreign entities in AI (see Appendix G).

There are, nonetheless, various government initiatives aiming to advance the principle of inclusive and sustainable growth, including, in particular, those of the National Academy of Science and Technology (NAST) and the National Economic and Development Authority (NEDA). NAST's Pagtanaw 2050 is a multidisciplinary and stakeholder-driven plan on the national development of science and technology. NEDA's AmBisyon 2040 is a national economic objective that champions well-being and improvement of quality of life. AI development is a natural extension of these initiatives. Comparing countries with AI national frameworks (see Appendix A), inclusiveness is a recurring theme, even as countries like China, Korea, UK, and the US only briefly mention this principle. In particular, Korea connects inclusive growth with job-safety (OECD.AI, 2019).

2.2. Human-centered values and fairness. Throughout its lifecycle, an AI system has to be human-centric and fair, and fairness is essential to treating people with dignity and respect . AI actors have to treat people fairly, avoiding algorithmic decisions and their consequences that are discriminatory. Fairness entails an AI system's respect for human rights and people's **data privacy rights**. AI actors should not implement mechanisms that tend to undermine the human capacity for self-determination. Safeguards "appropriate to the context [of AI application] and consistent with the state of art" have to be put in place in order to maintain fairness (OECD, 2019).

Human-centered values necessitates the protection of human rights. According to the Commission of Human Rights (CHR), AI will have a significant impact on political rights

(see Appendix D). Various sectors may be vulnerable to manipulation through the utilization of AI. An analogous local example would be the Cambridge Analytica controversy of Facebook involving data of Filipino users exploited for the elections (Gutierrez, 2018). Another particular concern is the red-tagging of government critics through the use of social media (Human Rights Watch, 2017). AI could potentially amplify such practice. Thus, CHR proposes that respecting human-centered values comes with more legal protection of human rights, adjusted to the new challenges that AI brings.

Defining "fairness" is of particular concern to AI startups and representatives from the academe. A clear, unambiguous definition of fairness is essential, at least, to compliance. AI systems are largely dependent on available data that may contain latent bias. Algorithms could also be driven by models that are unintentionally biased. Thus, in evaluating fairness of an AI system, both data and algorithms must be assessed for bias (Personal Data Protection Commission Singapore, 2020). An ISO Standards for AI (ISO/IEC TR 24027:2022 for defining 'fairness' and addressing bias) has been suggested in this regard.

To illustrate the point of fairness, consider gender bias in AI. Only 22 percent of AI and data science professionals are women, resulting in a "feedback loop shaping gender bias in AI and machine learning systems" (Young et al., 2021). The impact of such lack of diversity in general and, in particular, women's participation in tech and AI cannot be underestimated. Already such deficits have been implicated in wrongful and discriminatory algorithmic decision making (ADM). For instance, Apple was once caught giving a husband 2 orders of magnitude greater creditworthiness than his wife, even if both earn about the same income and have about the same spending levels and credit limits (Smith & Rustagi, 2021). This kind of discrimination is not isolated. Gender bias has been observed in other areas of AI application (such as job recruitment, profiling, facial recognition).

Two sources of bias are identified from the consultations: (1) data fed into the AI system and the (2) model being utilized. In a Harvard Business Review article, the first source can be traced to "flawed data sampling, in which groups are over- or underrepresented in the training data". With the second, algorithms are also at risk based on their training data (Manyika et al., 2019). Further look into neural networks may also be conducted with its promising capability to overcome dataset biases (Zewe, 2022). Understanding such biases and their sources could lead to fairer AI systems.

2.3. Robustness, security and safety. An AI system has to be robust, performing reliably and safely. Security is essential to its operations (Microsoft, n.d.). Notably European national AI strategies seek a 'robust AI ecosystem' (Van Roy et al., 2021) as safeguard against possible errors in processing and outcomes that could lead to harm, especially

involving applications like self-driving cars, robotic surgical assistance, control of electrical power grids, and the like (Dietterich, 2017).

Representatives from select Philippine government agencies emphasized the role of regulation in upholding this principle. Bangko Sentral ng Pilipinas (BSP), for instance, shared that data masking tools are being utilized to maintain security and protect data privacy (see Appendix D). CHR representatives also maintained that regulations help ensure security and safety and the protection of human rights and avoid malpractices and prohibited uses of AI (see also Section 4 below).

A robust well-maintained cloud infrastructure can help boost secure and safe AI computing. In this regard, consultations with Microsoft and other businesses also highlighted the importance of adopting a Cloud-first Policy for infrastructure rollout. With cloud computing, the overall security and safety of AI systems can be pursued through appropriate policies and controls, without sacrificing cost-efficiency (Jose, 2022).

2.4. Accountability. The proper functioning of an AI system depends on AI actors assuming responsibility and accountability in their roles, environments, or contexts. Accountability has at least three elements (Loi & Spielkamp, 2021):

(i) *Responsibility* for actions, providing grounds for moral praise or blame, social approval or liability to (possible) legal sanctions;
(ii) *Answerability*, which entails:
   a. Capacity and willingness to share the reasoning involved in the decisions made by the AI system, and
   b. entitlement of regulators or auditors to request for such reasoning; and
(iii) *Sanctionability* of the determined accountable actor or party.

Answerability implies **auditability**, allowing, for instance, third parties (especially regulators) "to probe, understand, and review the behaviour of the algorithm through disclosure of information that enables monitoring, checking or criticism" (Personal Data Protection Commission Singapore, 2020). With accountability being closely linked to "answerability", explanation or justification becomes important (Busuioc, 2020).

Auditability in AI may require an automation auditor outside the accountable organization. Such an auditor would have some kind of legal responsibility to ask certain questions and receive truthful answers, indirectly giving the public some level of control over the processes and decisions undertaken by an AI system (Loi & Spielkamp, 2021). This helps ensure that individuals and organizations deploying AI technologies are accountable for their consequences and are required to periodically update precautionary measures for mitigating associated AI risks (Amazon Web Services, 2021).

Some stakeholders claimed that current AI systems are “black boxes” that pose a serious challenge to the principle of accountability (Appendix D). But do these systems in fact escape human comprehension and therefore their accountability is hardly going to be associated with specific humans? AI systems are not ‘indecipherable black boxes’, and that explanation can be demanded from them, maintains Doshi-Velez et al. (2018). These are explicable, and explainability enables accountability (Floridi, 2021).

Other stakeholders suggest the establishment of an inter-agency body to provide oversight for AI and ensure accountability. This body may receive reports and recommend penalties for those who use AI for crimes, discrimination, and violation of human rights. To ensure proper compliance and auditability, the inclusion of appropriate agencies would be vital (Maharjan, 2022).

2.5. Transparency and explainability. Transparency of policies, rules and regulations governing AI systems is indispensable to a democracy. In the proposed AI regulation of the European Commission (2021), transparency is an obligation for AI systems

(i) “interact[ing] with humans”
(ii) “used to detect emotions or determine association with (social) categories based on biometric data”, or
(iii) “generat[ing] or manipulat[ing] content (‘deep fakes’)”.

Microsoft has also identified five parts of the funnel that need to be addressed under the principle of Transparency (Jose, 2022), namely:

(i) Full disclosure about the use of an AI system in decision-making
(ii) Intended purpose of the AI system
(iii) Training data
(iv) Maintenance and assessment, and
(v) Ability to challenge and seek redress.

Strong transparency is needed in how AI technologies are used in the public sphere and in commercial applications (Amazon Web Services, 2021). Such a transparency rule allows end users of AI, or people who may be affected by it, to make informed decisions, enabling them to opt out or “step back from a given situation” (European Commission, 20021). Transparency helps provide safeguards to AI systems and promote individual autonomy (Loi & Spielkamp, 2021).

Transparency requires that an AI system (or at least the essence of it) is *understandable*, even to non-technical stakeholders. The Singapore framework also required *explainability*, ensuring “that automated and algorithmic decisions and any associated data driving those decisions can be explained to end-users and other stakeholders in non-technical terms” (Personal Data Protection Commission Singapore, 2020).

Transparency also encourages informed public debate, helping AI acquire a certain level of democratic legitimacy (Loi & Spielkamp, 2021).

However, the concern of AI systems being “black boxes" that could undermine transparency and explainability has been repeatedly brought up in stakeholder consultations. But even some stakeholders would offer a broad AI education to help promote widespread understanding of AI, hopefully driving AI systems to be more transparent in the eyes of the public. Education promotes ‘computational and data literacy’ which could help address the concern of AI as blackbox systems. But the flipside to transparency could be that information abundance is potentially overwhelming to users and the public, thus rendering AI systems even more obscure (Floridi, 2021).

Transparency entails *engagement* of stakeholders. To whom do AI systems have to be transparent? If not in the context of governance and engagement, why be transparent in the first place? In being transparent, accurate and complete logs must be maintained, allowing proponents of AI systems, for instance, to “articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst-case implications can be understood and can inform mitigation procedures” (Personal Data Protection Commission Singapore, 2020).

Some stakeholders have recommended qualifiers on the principle of transparency (see Appendix D). One is for its limited application to policies and processes and not for AI algorithms themselves. Applying transparency to all elements of AI poses certain risks (Floridi (2021), including transparency becoming “detrimental to innovation”. Too much transparency “unnecessarily divert(s) resources that could instead be used to improving safety, performance and accuracy” (Floridi, 2021).

Furthermore, in her BBC Reith Lectures two decades ago, Prof. Onora O'Neill (2002) observed that the enthusiasm for more transparency and openness “has done little to build public trust. If anything, trust has receded as transparency has advanced.” Today her observation continues to convince.

2.6. Trust. Trust is a distinct principle and is hardly a condition resulting from transparency. It is often cited as a key desirable property of the interaction between any user and AI system (Jacovi, 2021).

Trust is an important predictor of a society’s willingness to adopt a range of AI systems — from AI-enabled banking systems to autonomous vehicles, from product recommendation agents to medicine dispensing robots, and so on. However, the current developments in AI are not helping build trust, so far. The European Commission’s AI High-Level Expert Group (AI HLEG) notes that if AI systems do not prove to be trustworthy, their widespread acceptance and adoption will be hindered,

and the vast potential of socioeconomic benefits will remain unrealized (Lockey et al., 2021).

Essentially the principle of trust involves people's willingness to be vulnerable based on "good reasons" (Lockey et al., 2021). Without trust, mutually beneficial transactions and agreements are at best costly, if not hard to come by. As former Secretary of State George Shultz (2020) puts it, "Trust is the coin of the realm. When trust was in the room, whatever room that was—the family room, the schoolroom, the coach's room, the office room, the government room, or the military room—good things happened. When trust was not in the room, good things did not happen. Everything else is details."

With trust, people are willing to experience whether AI systems can be operated safely, reliably, and consistently even under difficult (if not unexpected) conditions, especially for applications (say, in healthcare, financial services, transportation and the like) affecting lives and livelihoods and involving consequential decisions (Jose, 2022). As Dr. Lobana puts it, "In order for the general public to trust AI-based products or services, they need to have faith that there are regulations that are ensuring that AI products or services are not released in the market until they meet a certain threshold of safety, security, robustness, fairness, transparency, and other key required attributes. For instance, the general public trusts the use of electricity and electrical products as they know that the producers of the electrical products need to ensure that certain safety thresholds are met before they can release the products in the market" (Correspondence, 27 January 2022).

Trust is earned, not always readily given. Before employing a nationwide adoption of AI systems, five central AI *trust challenges* should first be tackled, namely (Lockey et al., 2021):

> (i) *Transparency and Explainability.* AI systems that do well are not necessarily the most transparent and explainable. On the other hand, the more comprehensible ones are not necessarily the most accurate or reliable;
>
> (ii) *Accuracy and Reliability.* Inaccurate outcomes of AI processes can lead to bias, unfairness, and potential harm to end-users and groups;
>
> (iii) *Automation.* Decisions that implicate social and ethical norms tend to defy automation. Perhaps the next best thing is *augmentation*, allowing human collaboration with machines to perform a task. Some stakeholders affirmed augmentation by incorporating a Human in the Loop (HITL) process (see Appendix D);

(iv) *Anthropomorphism*. Most current AI systems are not (yet) human-like. But over-anthropomorphism may lead to overestimation of an AI system's capabilities, which could potentially put users and groups at risk;

(v) *Massive Data Extraction*. The extraction of massive amounts of data for the development and implementation of AI systems poses serious privacy risks. For many users and stakeholders, loss of privacy and unwarranted sharing of information is problematic.

As AI systems earn trust, demonstrably they become more trustworthy. Beyond these systems being able to reliably perform or fulfill their designated and expected purpose (Millar et al., 2018), some representatives from Philippine government agencies also emphasized how contributory to AI's trustworthiness are the adherence to the standards of competition laws and the very AI governance framework being inclusive enough to encourage fair competition among AI businesses and AI-powered organizations (see Appendix D).

For a sense on how these principles have been deployed in the various national strategies, plans, frameworks of the different countries, see Appendix A.

3. Elements of AI Governance

Lobana (2021) has done a comprehensive accounting of diverse IT governance frameworks out there. The following table shows selected elements that we have considered of key importance for our research:

Figure 1. Comprehensive accounting, lexical ordering of AI elements

Governance Element	Description
1. Board Oversight	An overseeing board handling all IT processes of the organization.
2. Communication Tools	Exchange of necessary information among the organization's stakeholders through the use of communication processes.
3. Culture, Ethics & Behaviors	The organization's culture, ethics, and individual behaviors
4. Human Behavior	Consideration of how IT-related processes impact humans (employees, customers, etc).

Governance Element	Description
5. Information	Relevant information produced by or used by the enterprise that is applicable to the effective functioning of the enterprise's governance system.
6. Investment Priorities	IT investment decisions and priorities for funding AI projects and initiatives
7. IT Architecture	Consideration of the organization processes' architectural maps, data, applications and technology; and assistance in IT strategy formation and integration and standardization across organization.
8. IT Audit	Audit of IT-related systems and processes
9. IT Coordination	Coordination of IT activities in support of organization's strategic objectives
10. IT Governance Framework	An overall framework for IT governance in an organization.
11. IT Principles, Policies, and Procedures	Setting up principles to guide the IT activities within an organization. Providing detailed guidance to IT executives and personnel through IT Policies and Procedures.
12. IT Steering Committee	An executive committee that oversees an organization's IT governance, comprising representatives from both the business and the IT departments.
13. IT Strategy/Governance Committee	IT governance is overseen by a board-level committee.
14. Organizational Structures	Roles and responsibilities for IT decision-making, as well as related accountability as parts of organizational structures.
15. People, Skills, & Competencies	Readily available human resources with the necessary skills and competencies to meet the key IT goals.
16. Performance Management	Processes to manage the performance of IT projects, investments, and other resources

Governance Element	Description
17. Regulatory Compliance	Processes to comply with rules and regulations in jurisdictions where the organization does business
18. Resource Management	Processes to manage the organization's resources most efficiently and effectively
19. Risk Management	Processes to manage IT-related risks, including information security risks as well as operational and systemic risks arising from the use of IT
20. Services, Infrastructure & Applications	To aid in the delivery of the primary IT objectives, services, infrastructure, and applications are available.
21. Stakeholder Management	Processes to manage various stakeholder relationships inside and outside the organization
22. Strategic Alignment of IT	Enhancing the alignment between business and IT through setting up of management processes
23. Strategic Planning	Long term planning in line with overall organizational strategy
24. Value Delivery	Maximization of returns from IT investments and optimization of expenses

Figure 1 presents a lexical ordering of the selected elements from Lobana's accounting of IT governance elements. Initial consultations with local stakeholders indicate that one organization may emphasize one or two elements more than the others. While it may be too early in the stakeholder consultation process to say which elements have greater salience in Philippine organizations, by far, local AI startups have identified performance management, regulatory compliance, and strategic planning as their top AI governance elements to consider (Department of Trade and Industry, 2021, 21 December).

Widely used approaches to the framing or organizing of these elements have been codified by ISACA (Cobit 2019) and the ISO (ISO 38502, 2017). Invariably none of the approaches would use *all* the elements cataloged by the Lobana survey.

In addition, one can raise a question of distinction between IT governance and AI governance. Is the latter reducible to the former? Or are they arguably distinct, despite having many overlaps? Lobana's AI Governance Framework identifies key governance areas as an organizing mechanism for important governance elements (Figure 2).

Figure 2. Lobana's AI Governance Framework (Lobana, 2021)

Governance Area	Governance Elements
Engaged Board Oversight	Knowledgeable Board
	Engaged board
Enterprise Leadership & Planning	Competent, Committed, & Collaborative Top Management
	Focused AI Strategy & Risk Capital
	Enterprise Architecture & Coordination
Core AI Technical Elements	Governance of Data Assets
	Governance of Algorithms and AI Models
	Infrastructure Scalability
People & Culture	Strategic People Governance
	Culture of Innovation
	Change Management & Communication
Operational Structures, Processes & Mechanisms	Redesigned Processes
	Operational Structures, Policies & Practices
	Performance Management
	Stakeholder Management
Enterprise Risk Oversight	Risk Management & Audit
	Data & AI Security
	Regulatory Compliance
AI Ethics	Embedded AI Ethics
	Corporate Social Responsibility
Ongoing Evolution	Continuous Digital Transformation
	Evolving Holistic System

In this framework, an engaged board oversight is a governance area that covers overlapping elements of knowledgeable and engaged board. An organization's knowledgeable board would be responsible for overseeing the entire AI affairs and knowledge related to AI risks and opportunities.The more immediate need, however, is to address the lack of expertise in this area. Training board members towards this end, especially in a country where AI is fairly new, is a must to have a respectable AI governance. Focused on business strategy affected by AI, on the other hand, is the engaged board.

For AI to work in any organization, good leadership and management are essential. So are AI-driven strategy and architecture processes (Lobana, 2021). Future proofing your organization, however, would require more than just leadership and management.

Competent and talented AI specialists are not only good to be around with but they may have to collaborate with people outside the organization.

Organizations must have a robust governance of data assets, including their sourcing, monetization, standardization, security, and regulatory compliance. Considered more valuable than the governance of algorithms, data assets governance is the building block for the success of AI. In the governance of algorithms and AI models, boards must also ensure meticulous validation tests, as most of the algorithms used are obtained from diverse open-source libraries with uneven quality (Lobana, 2021).

With opportunities that AI brings come new multifaceted risks that the organization has to deal with (see Sec 7 "Risk-based Governance and Regulation" below). No organization can navigate through these risks and make sound decisions without effective governance and a clear moral compass. In particular, dealing with AI means dealing with "data and individual privacy, bias and fairness, trustworthiness, safety, robustness, and explainability" (Lobana, 2021). Conducting meaningful stakeholder or shareholder consultations to discuss many ethical concerns relating to AI is not only about good governance but also part of risk management. AI governance necessarily involves an examination of organizational goals and decisions as to which risks the organization is willing to take. Stakeholder management is an essential part of AI governance.

4. Humans and AI

4.1. Human-centricity for outcomes and processes. "AI solutions should be human-centered" (Personal Data Protection Commission Singapore, 2020). Important AI systems are meant to increase productivity, creating optimal or efficient solutions to problems humans face in their daily lives. Various stakeholders consulted have emphasized the same value, in the light of AI's potential to displace human workers.

While AI "has the potential to transform our world for the better", it is "not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being" (European Commission, 2019). Although AI systems are rapidly developing which threaten human resources, the World Economic Forum (2021) among others has developed a practical toolkit for human resource professionals.

Improvement of HR processes includes identifying key areas of concern through the adoption of AI augmented tools for human resource assessment teams while evaluating and monitoring the risks associated with such tools (World Economic Forum, 2021). Job security can also be aspired for. Part of Korea's National AI Strategy, for instance, is 'establishing an inclusive job-safety network' (OECD.AI, 2019: 44).

As AI continues to create new solutions to problems, human involvement is necessary as well to ensure that the systems work to amplify human capabilities rather than undermine them and cause harm to individuals and society. As AI opens countless doors to innovation, it does the same for potential misuse. While every outcome cannot be fully guaranteed to be fair in the deployment of an AI system, argues one representative in a stakeholders meeting (Department of Trade and Industry, 2021, 21 December), its "end-users need to be able to know how an AI comes to a conclusion, and thus how one can change the result." A human "needs to be in the loop" in an otherwise purely algorithmic decision making (ADM), making human accountability and system auditability clear and unambiguous.

By definition, however, "an autonomous system or function is to some degree out of human control" (International Committee of The Red Cross, 2019). So, human involvement in AI systems needs qualifications, especially in relation to spheres of influence or areas of AI application. High-risk applications of AI would specially need "humans in the loop" (HITL). Full Machine "autonomy" cannot exist in a vacuum. It requires support infrastures and systems. These, in turn, are dependent on the country's level of economic and technological development (see Appendix G).

Human involvement in AI is not limited to HITL, leveraging both human and machine competences in a virtuous cycle that produces value and good outcomes. At the core, AI should be protective of human rights (HR).

Ultimately AI systems are not rights-neutral. Some AI applications tend to undermine HR; others tend to support (Raso et al., 2018). Good practices supportive of HR should be highlighted. For instance, an initiative of UPR Info to collect and categorize United Nations recommendations on improving human rights to 'pressure' each state to meet their HR obligations is largely dependent on AI tools (Finch, 2020; Saslow and Lorenz, 2019). The pressure to implement a human rights-based governance facilitates both adherence to AI principles and the promotion of rights itself. Using AI, regulatory models of the European Union and the Council of Europe may be used to compare and determine how governance can protect human rights (Cataleta & Cataleta, 2020). Human-centricity is sustained through the protection of human rights and active avoidance of prohibited uses of AI.

4.2. Prohibited Use of AI. Ultimately, AI governance or regulation is all about human accountability. Determining the structures or parameters under which AI systems can operate is essential (Sanford, 2021; Doshi-Velez et al, 2018). Well-defined roles and responsibilities of humans in AI can be codified regulations and professional standards. Clarity of what is allowed and what is prohibited can be determined in advance.

4.2.1. Lethal Autonomous Weapons Systems or "killer robots". Such machines have to be banned. A High Contracting Party to the Convention on Certain Conventional

Weapons (aka the “Inhumane Weapons Convention”), the Philippines has to support the Stop Killer Robots Campaign (n.d.). The Convention is a framework for banning or restricting weapons considered to cause unnecessary, unjustifiable and indiscriminate suffering (United Nations, n.d.), As one of the local AI startups argues, at least we can start with what AI should not be, and killing or harming people must be one of them. Killer robots are lethal ADMs that make decisions with little or no human involvement (Walsh, 2021).

In addition, the European Commission (2021) is proposing to prohibit practices deemed to pose risks to individuals, including manipulative and exploitative use of AI, indiscriminate surveillance, and social scoring.

4.2.2. Manipulative and exploitative practices. Some AI systems “have a significant potential to manipulate persons through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm” (European Commission, 2021). Targeting individual vulnerabilities through choice architecture or user-interface (UI) manipulation could soon be regulated in the EU territories.

Manipulative and exploitative practices in AI can have broad impact, not the least of which is the undermining of privacy. “Privacy is the…most impacted [right] by current implementations of AI” (Raso et al., 2018). To illustrate, Cambridge Analytica facilitated the manipulation of Filipino voters in the 2016 Presidential Elections through and with the help of AI-powered Facebook (Occeñola, 2019). In a similar vein, AI-powered platforms have facilitated the circulation and consumption of pornographic contents (especially involving women and children) that violate data privacy (Brutas, 2017). The UK Center for Data Ethics and Innovation (2020) reports that the ‘erosion of privacy’ as high risk for criminal justice and medium risk for digital & social media. Violations of privacy are violations of human rights.

4.2.3. Indiscriminate surveillance. The proposal to prohibit indiscriminate surveillance includes the prohibition against “the use of ‘real time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement” (European Commission, 2021).

Clearview AI is a surveillance platform with great potential for abuse. It utilizes facial recognition, allowing images to be linked to other photos of an identified individual. This kind of technology is especially concerning in Russia where AI is reportedly used for political targeting (Deutsche Welle, 2021; Human Rights Watch, 2021). Both the USA and China are in a virtual arms race, deploying AI-driven mass surveillance using biometric databases (Mint Press News, 2020). China, in particular, has expanded its surveillance systems and capabilities by incorporating biometrics, big data analytics,

and security softwares and exporting them to other countries (Li, 2020). The lack of regulation in this area predisposes these systems to more potential for abuse. However, large-scale tracking of individuals in virtual and physical environments using AI could soon be regulated in Europe.

Raised by the CHR during stakeholders consultations are some local examples of indiscriminate surveillance that led to harassment of Philippine government critics, especially after the enactment of the Anti-Terror Law (R.A. 11479). Making indiscriminate surveillance possible, such law undermines human rights (Amnesty International, 2020; International Service for Human Rights, 2020; Foundation for Media Alternatives, 2016). Such kind of state surveillance has been used to silence dissenters, threatening not only their human rights but also their lives (Pacific Media Watch, 2018). Certainly AI can boost existing surveillance programs and help inaugurate new ones.

4.2.4. Social scoring. AI-based social scoring of natural persons by public authorities should also be prohibited (European Commission, 2021). It "may lead to discriminatory outcomes and the exclusion of certain groups. They may violate the right to dignity and non-discrimination and the values of equality and justice" (European Commission, 2021). But as BSP (Banko Sentral ng Pilipinas) representatives pointed out in a stakeholder consultation meeting, those in the lower classes of society are most susceptible to injustices that AI systems may inadvertently cause (Appendix G).

Overall, AI should not be used "for manipulative, exploitative and social control practices" (European Commission, 2021).

5. Data Management

Data fuels AI. With the emergence of AI systems and increasing utilization of AI-augmented products and services, the need for reliable, comprehensive, and instantaneous data becomes more pressing. As systems are established to acquire more and more data, certain ethical and governance concerns arise. Data governance issues, such as those relating to consent, ownership, and privacy, can be aggravated by improper utilization of data for AI (Floridi, 2021). Steinhardt and Toner (2020) noted that while policy interventions strengthen AI systems, they could also weaken these systems. For instance, in the context of the COVID-19 pandemic, AI grappled with the continuous change of data and produced inaccurate results (Steinhardt and Toner, 2020).

Data also has a life cycle of its own (from acquisition to retirement or disposal). An overall data management strategy is as important as putting someone accountable for the management process itself. The initial acquisition of data is critical as it may come in a variety of forms. Data collected has to be cleansed and processed before getting integrated into AI systems. Organizations must specify data quality standards that they

wish to maintain. They have to implement mechanisms to ensure data is of optimal quality to produce trustworthy results (Lobana, 2021).

On the macro level, the Philippine AI Roadmap points to the need for massive amounts of data for the development of AI systems, to "make public data open, freely available, and downloadable in digestible format..." (Task 6). The UK National AI Strategy refers to data that is FAIR (findable, accessible, interoperable, reusable), forming part of the so-called "data foundations" or characteristics that "contribute to its overall condition, and ensuring [that] it is fit for purpose, and recorded in standardised formats on modern, future-proof systems" (UK Office for Artificial Intelligence, 2021). Organizations reporting higher levels of AI adoption have excellent data foundations. Data foundations are a necessary (but insufficient) condition for the adoption of AI (Ernst & Young, 2021).

5.1. Open Data. Towards the goal of making national public data FAIR, the Open Data Philippines (ODPH, n.d.), a government program collecting datasets from different government agencies, needs more headway, attuning to the data needs of data scientists, AI researchers, policy makers, and other data stakeholders. One of the major initiatives under such a program is the Philippine Statistics Authority's (PSA) OpenSTAT, allowing the PSA "to share data under an open data license where data can be freely used, re-used and redistributed by anyone without any restrictions other than proper source attribution" (Philippine Statistical Authority, n.d.). OpenSTAT's objectives could be well be an integral part of data management and governance policy in general and, in particular, for AI development, including:

- "adherence to the Open Data Initiative"
- contributing "towards [the] achievement of Sustainable Development Goals"
- facilitating "an inclusive, sustainable and resilient development"
- promoting "a National Data Sharing, Accessibility Policy and Standards"
- promoting "innovation through provision of Open Application Program Interfaces"
- increased and improved "utilization of data for decision-making, citizen empowerment, innovation and entrepreneurship"
- supporting "capacity building and innovation for the generation, sharing and utilization of data at national, regional and local level"

5.2. Data as source of bias. A fair and gender-smart data management will have to address gender bias and other forms of prejudices that may creep in at every stage of the data lifecycle. Bias leading to discrimination is one of the common risks evident in criminal justice, financial services, health and social care, and digital and social media (UK Center for Data and Innovation, 2020). To quote Raso et al. (2018, p. 18), "AI systems are trained to replicate patterns of decision-making they learn from training data that reflects the social status quo—existing human biases, entrenched power dynamics and

all." Sound data management has to systematically address biases potentially latent in data sets.

Three processes ensure fairness: data quality management, metadata management, and data access management (Harrison et al., 2019). **Data Quality Management** "consists of choosing quality dimensions that are appropriate for assessing the suitability for use of a dataset based on what is critical for business operation, reporting, or other relevant purpose" (Harrison et al., 2019). To try to free data of bias and help ensure non-biased outcomes, data must be filtered and vetted. **Metadata Management** is the process of "planning, implementation, and control activities to enable access to high quality, integrated metadata" that ensures the robustness of data (Harrison et al., 2019). Lastly, **Data Access Management** "refers to the principles by which security policies and procedures are defined, planned, developed and executed" (Harrison et al., 2019, 2.3). These interacting processes are essential to ensuring biased-free data and outcomes.

Data management is a core competency, a "prerequisite to the establishment of robust AI initiatives" needing data governance (Harrison et al., 2019). The function of data governance is to formulate decisions and policies that identify the expected actions of people and processes in accordance with the data (Harrison et al., 2019). Sound data management and governance help build trust in AI.

6. Internal and External Governance Structures

Frameworks are instruments that enable organizations to operationalize otherwise abstract principles and values, incorporating them into actual plans, policy choices, approaches, strategies, procedures. Here "organizations" is taken to mean broadly to include both private and government organizations. Arguably even the whole of society is an organization. Without internal organizational governance structures and measures for day-to-day operations and decisions, such principles and values are empty.

An internal governance structure has to be able to support the following roles and responsibilities in the deployment of an AI system (Personal Data Protection Commission Singapore, 2020): oversight, training, and risk management. Additionally, the engagement of internal and external stakeholders has to be an integral part of an organization's structures and functions.

6.1. Oversight. The overall responsibility of providing oversight at every stage of the governance framework is clearly defined and assigned to appropriate personnel, teams, or departments. An effective oversight mechanism requires coordination with relevant experts, representatives or stakeholders across the nation. On the macro-level, the government is in charge of overseeing the AI ecosystem and building a standard from

which the whole system must follow (Lobana's AI Framework, Figure 3 above). Stakeholders consulted also favor the establishment of a body for AI governance and oversight (see Appendix E).

6.2. Training. The organization needs to ensure that personnel given specific responsibilities have the requisite training for their respective roles. A robust AI ecosystem requires a competent workforce developed through training. Stakeholders from diverse sectors consistently point training as an essential factor in AI development. The UK's national AI strategy also emphasized the importance of expanding the skills and talents of AI developers and even users (Section 2.3). For this to materialize, the academe plays a huge role in providing training in AI. Appropriate government agencies need to specify the standards against which training competencies are measured to produce competent AI practitioners.

In part, Section 4 above discusses the threat of worker displacement by AI, and therefore measures (regulatory or otherwise) to boost workforce upskilling and reskilling have to be initiated immediately to help provide social protection for workers (Appendix D; Amazon Web Services, 2021). As the Federation of Free Workers also maintains, the upskilling and reskilling of the Filipino workforce entails making AI a job-creating system for workers (Appendix D). Korea's national AI strategy regarding job-safety is a good reference for promoting social protection.

6.3. Risk Management. Risk management involves a range of options including risk transfer and risk avoidance. But inevitably an organization has to exercise internal risk controls. The nation is expected to adopt or develop its own effective risk management framework that serves as guide for both the management and the rank-and-file. Key considerations of such key management frameworks include:

(a) Assessment and management of risks and potential adverse impact on vulnerable individuals and groups;
(b) Review and monitoring of AI models deployed;
(c) Review and monitoring of communications or interactions with stakeholders;
(d) Review and monitoring of personnel actions on identified risks.

Diverse risks are properly discussed in greater detail using risk management frameworks and standards for organizations (e.g., ISO 31000, NIST 800, etc.). But salient in the minds of local stakeholders are risks associated with lack of data sharing for the training of AI, data privacy and security breaches, and "biased" AI. One of the AI startups is specially concerned about local AI companies being vulnerable to "brain drain" of local AI talent going abroad. The UK, for instance, is set to "roll out new visa regimes to attract the world's best AI talents" (UK Office for Artificial Intelligence, 2021).

There are AI risks that go beyond the purview of separate, individual organizations or even governments. When AI systems get connected, for instance, beyond national borders, it behooves the community of nations to cooperate and set an international framework of AI governance. Hence, there is a need to establish structures external to individual organizations in order to deal with general parallel needs to exercise oversight, facilitate training requirements, and manage shared risks in the development of AI.

6.4. Stakeholder Engagement Processes as Structural Function. The Foundation for Media Alternatives claimed that a broad stakeholder-driven approach ensures the participation of various sectors in strengthening AI (Appendix D). Companies and AI-powered organizations have internal stakeholders that need to be consulted as part of risk management. External stakeholders will also have to be involved. Japan's Ministry of Economy, Trade and Industry (2021) deems it best practice to not limit stakeholding to companies but broadest sectoral engagement possible (including users, engineers, academics, and law/audit experts). Such a stakeholding process is "desirable" with the government functioning "as a facilitator in the discussion and objectively evaluat[ing] whether companies satisfy the guidelines developed…" (Ministry of Economy, Trade and Industry, 2021).

Beyond the immediate commercial interest of companies, a broad stakeholding process should include sectors and groups most likely to be negatively affected by AI. Bearing the risks of algorithmic systems in mind, these people should have a say on whether and how the algorithms are used, argues New York University's Meredith Whittaker (2020). For the Philippines, perhaps the extent of stakeholding necessary could be inferred from the EU process involving hundreds of businesses, civil society organizations, public authorities, experts, etc. (European Commission, 2021).

6.5 Professional Regulatory Body. Composed of entities or agencies outside AI-solutions providers and private organizations, a professional regulatory body would ensure user safety and legal compliance in the adoption of AI. Policies and external mechanisms such as government and professional regulations or guidelines for best practices would be drafted and enforced by this external professional regulatory body. The regulatory body would also coordinate policy development activities in national and local settings. It may examine an organization's chain of responsibility for all roles involved in the design and development of AI systems (Loi & Spielkamp, 2021). The body also needs to facilitate the development of codes of conduct and accountability mechanisms for licensed practitioners, accounting for their ability to have broad and rapid impacts on society through the development, procurement, deployment and use of AI systems (Millar, et al., 2018).

On behalf of the regulatory body, an automation auditor may also be deployed. Such an auditor is tasked to examine and audit a provider's set of codes and data used in the

development of models informing their AI systems and practices. The results of the audit have to be made available to both internal and external stakeholders (Loi & Spielkamp, 2021). Auditors would also have a "right to explanation" when algorithmic decisions are made. They have a right to request truthful information clearly explaining the algorithmic logic used to render a decision when an AI system uses personal data. Auditors should adhere to rules concerning individual consent to personal data use during the audit process. They themselves have to abide by the standards for audit and for ethically performing algorithmic impact assessments and quality assurance (Millar, et al., 2018).

7. Risk-based Governance and Regulation

Traditionally risk management is best embedded in structures and functions internal to organizations. But as designs and operations of technological, organizational, and social systems continue to be driven by artificial intelligence to unprecedented levels, an overall risk-based approach to governance and regulation has to be articulated as well. Governments appear to be confident that they have a good handle of their regulatory powers necessary to develop AI for the benefit of society "while not being excessively prescriptive" (European Commission, 2020). The extent of regulation especially of emergent technologies like AI should be proportional to risks they pose.

Below is a reproduction of "common risks" associated with AI, according to the UK Center for Data Ethics and Innovation (2020):

Figure 3. Common Risks of AI. Key: Higher (🔴), Medium (🟡), Lower (🟢) Risks

Common Risks	**Criminal Justice**	**Financial Services**	**Health & Social Care**	**Digital & Social Media**	**Energy & Utility**
Bias leading to discrimination	🔴	🔴	🔴	🔴	🟡
Lack of explainability	🔴	🔴	🟡	🔴	🔴
Regulator resourcing	🟡	🟡	🔴	🔴	🔴
Higher-impact cyberattacks	🟡	🔴	🔴	🟡	🔴
Failure of consent	🔴	🟡	🔴	🔴	🟡

Common Risks	Criminal Justice	Financial Services	Health & Social Care	Digital & Social Media	Energy & Utility
mechanisms					
Loss of trust in institutions	🟡	🟡	🔴	🟡	🔴
Lack of transparency	🔴	🟡	🟡	🟡	🟡
Unequal access to services	🟢	🔴	🔴	🟡	🟡
Effects of low digital/data maturity	🟡	🟡	🔴	🟢	🔴
Erosion of privacy	🔴	🟡	🟡	🟡	🟢
Platform and data monopolies	🟡	🟢	🟡	🟡	🔴
Excessive data retention	🔴	🟡	🟢	🟡	🟢
Low ‘human-in-the-loop’	🟡	🟡	🟡	🟡	🟢
Mis/disinformation	🟢	🟢	🔴	🔴	🟢
Loss of trust in AI	🟢	🟢	🟡	🟢	🔴
Undervaluation of public data	🟡	🟢	🔴	🟢	🟢
Low accuracy	🟡	🟢	🟡	🟡	🟢
Undermining professional judgment	🟢	🟢	🔴	🟢	🟢
Excessive trust in AI tools	🟡	🟢	🟡	🟢	🟢

Resonating with local stakeholders are AI common risks, including bias leading to discrimination, security breach, erosion of privacy, cyberattacks, loss of trust in institutions and AI, lack of transparency, unequal access to services, platform and data monopolies through vertical ownership, and mis/disinformation (see Appendix F).

Some of these risks have already been taken up in the previous sections (e.g., Section 2, 4, and 5 above). Others need further amplification. For instance, the risk of bias leading to discrimination, along with the lack of explainability, poses the highest risk across sectors (Figure 3). Such a bias persists due to the replication of existing sequences in decision-making based on the social status quo (Raso et al., 2018). This is addressed in part by developing and constantly updating different status quos. The risk of lack of explainability, on the other hand, remains a challenge even for AI systems that perform well but more so for those that do not (see Section 2 above). Both bias and explainability risks intersect with the risk of lack of transparency. In one stakeholder consultation, AI startups push for explainability as a desired value but not as a required AI element (see Appendix D).

The risks of security breach, erosion of privacy, and cyberattacks are intertwined and could lead to people losing trust in institutions handling data for AI systems. Some local stakeholders also raised concerns that the risk for mis/disinformation in the Philippine financial services is rather understated compared with the UK's as assessed by the UK Center for Data Ethics and Innovation (2020). AI, according to a Proctor & Gamble (P&G) representative, could possibly be an 'enabler' of more mis/disinformation. In developing AI in the Philippines, gaining people's trust is a protracted challenge, even as an emergent technology, AI systems have not been immune to cybersecurity breaches. Properly deployed and monitored, AI can also be a tool to combat disinformation (Kertysova, 2018).

These risks common to criminal justice, financial services, health and social care, digital and social Media, and energy and utilities, are at best indicative of breadth and depth of AI's impact in society. The discussion of such risks here is not meant to be exhaustive. For lack of space, reflexive risks of AI could only be hinted at here. Vertical ownership of AI platforms and data monopolies, for instance, could be an environmental risk to AI's own development. In our government's desire to accelerate AI, unfair advantages and incentives could be given to certain companies which could in turn help promote anti-competitive practices. These "favored" companies could in effect be perversely incentivized to restrict rather than promote data sharing. So rather than develop AI in the long haul, initial regulations meant to champion AI in the short term could actually stunt technological growth, making the country a net loser. So, an argument can be made for the government to broadly and systematically address AI risks in order to lessen unintended negative consequences. In this light, stakeholders from the academe, for instance, propose to privilege the production of human capital for AI over narrowly

conceived AI initiatives. Qualified graduates across disciplines would be the essential workforce for a robust AI ecosystem.

Overall, in dealing with risks associated with the widespread deployment of AI, an equally *risk-based governance and regulation* need to ensure organizational and societal controls are clearly identified to mitigate the potential impact of such risks.

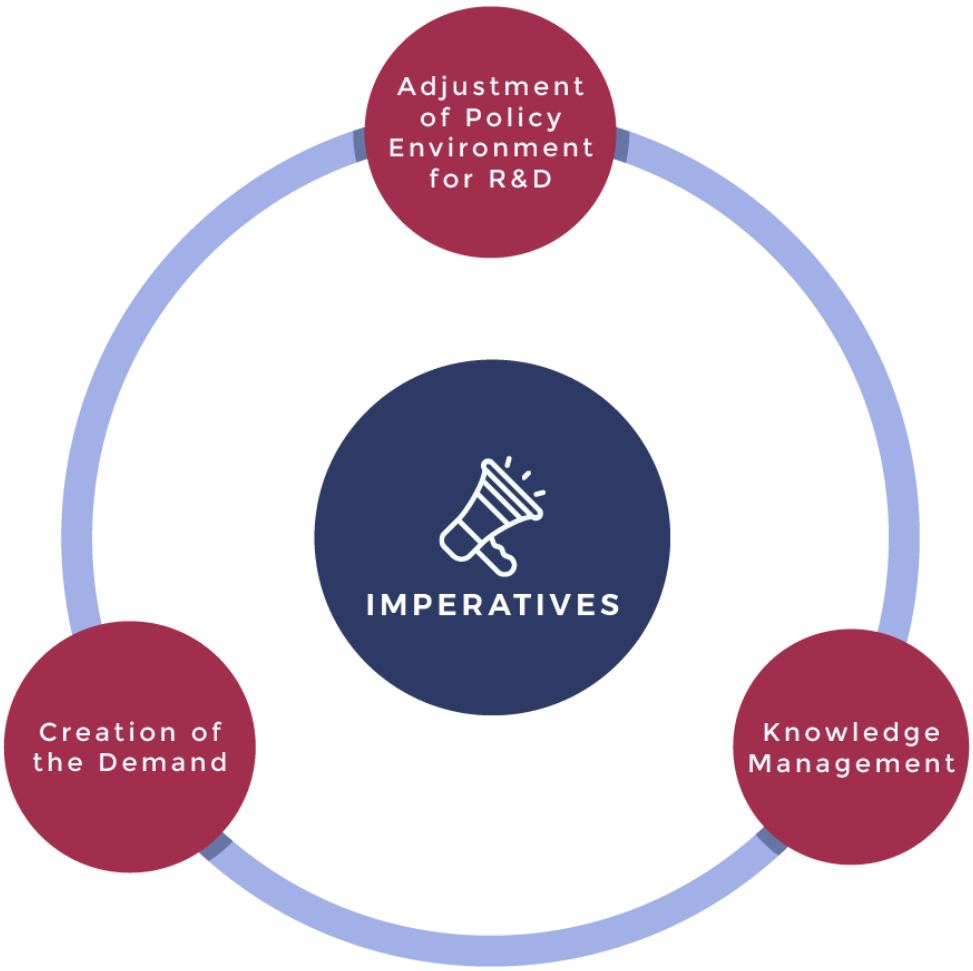
8. Inhibitors and Drivers of AI Development

A country's development of AI is driven or inhibited by various factors beyond individual actors and markets. A good AI governance framework cannot be developed without taking stock of an AI development's general drivers and inhibitors.

DICT Usec. Maria Castro identified three (3) imperatives to drive AI in the Philippines, namely (Fig 4):

(i) adjustment of the policy environment for research and development,
(ii) creation of demand, and
(iii) knowledge management.

Figure 4. Imperatives for AI Development (Castro, 2022)



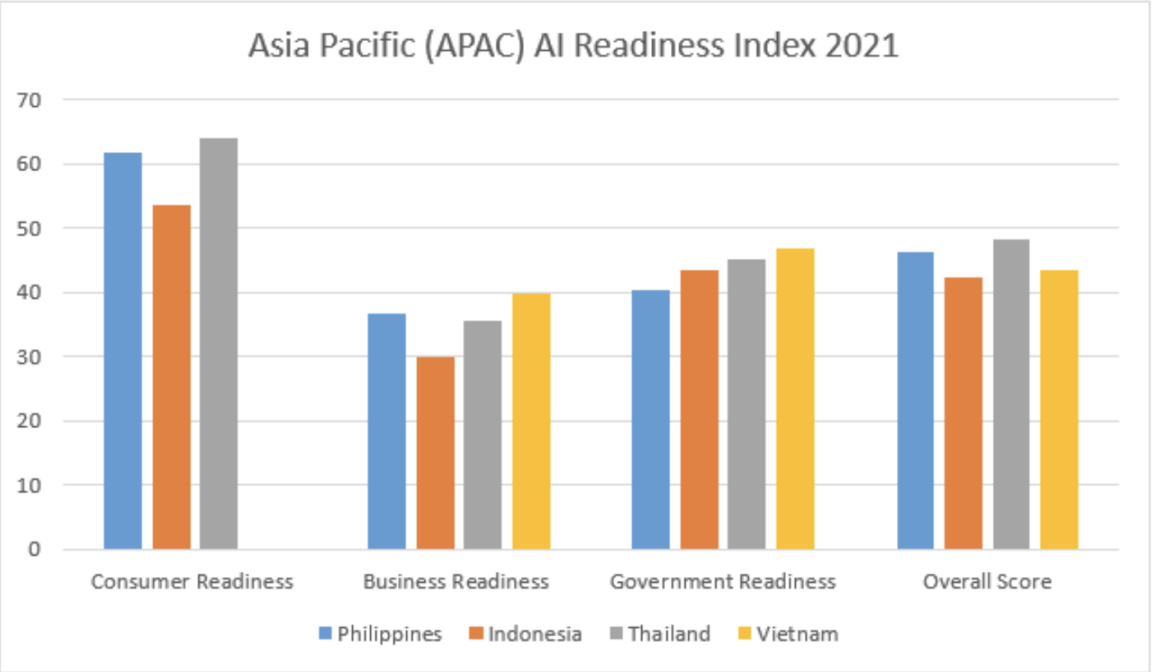
Technical innovativeness is essential to propel an economy and increase the competitiveness of its industries. Without strong and robust R&D programs, the country's long term progress and societal growth cannot be achieved. Hence, R&D and innovation initiatives such as AI must be placed at the center of major efforts to ensure that key industries remain adaptable to the ever-changing economic landscape and capable of embracing new economic realities. Moreover, R&D programs uplift the technical capacities of industries, enabling them to adapt emerging technologies such as AI. The capacity of AI-solutions providers and the technology sector is an important determinant of technological diffusion primarily because the initial conceptualization of an emerging technology like AI needs appropriate technical capacities and skills to make it commercially available (Hall & Khan, 2002). Policies supportive of R&D need to be periodically calibrated to ensure a healthy policy environment and to accelerate AI innovations.

Market demand is another driver. This is formed when an increased utility from the new technology meets the user's current technological readiness (Hall & Khan, 2002). The Asia Pacific (APAC) AI Readiness Index measures the AI frameworks and ecosystems present in Southeast Asia, considering factors such as adoption, deployment, and

support of AI technologies. According to the 2021 APAC report, the Philippine consumer readiness for AI is 61.8 on a scale of 0-100. Overall, the country scored 46.2.

In 2019, the Philippines averaged 44.2 in its overall scores, indicating that despite the COVID-19 pandemic, overall, AI readiness and adoption in the country increased. In the recent 2021 APAC report, however, the Philippine government readiness (40.4) is behind its comparable ASEAN neighbors (Fig 5). Indonesia, Thailand, and Vietnam scored 43.4, 45.0, and 46.7, respectively. Business readiness score puts the Philippines (36.6) behind Vietnam (39.9). Philippines' consumer readiness (61.8) is behind Thailand (63.9), yet significantly ahead of Indonesia (53.5).

Figure 5. AI Readiness of some APAC countries



The Philippine public sector in both the national and local levels is already receptive to the use of AI. However, the private sector is still in need of a competitive workforce to keep up with the utilization and development of AI (Castro, 2022). There is also a need for quality education for AI and the upskilling of talents to better equip private organizations in developing an AI-driven work environment, products, and services. The increased demand for IT services like AI also could pave the way for rolling out faster and more extensive data networks and the establishment of necessary IT infrastructures that would meet the demand from smartphones, online content consumption, and utilization of AI solutions (Castro, 2022).

The third imperative necessary to drive AI development is knowledge management in terms of the data collected, organized, and used (Castro, 2022). With the advancement of machine learning, collection of large sets of data (personal or otherwise) has become

necessary to adequately train AI models. The data sets that each government agency holds must not remain siloed and instead be used to create a data ecosystem, allowing government agencies and AI-solutions providers to share data safely amongst each other in ways respectful of people's privacy.

In the international landscape of AI development, Dirksen & Takahashi (2020) looked at "the relevant actors, the market, its policy ambitions, its challenges and of course opportunities for the Netherlands" as they studied the Japanese AI development and situation. They have identified Japan's political, economic, societal, technological, and legal drivers and inhibitors of AI development. While contextual differences would ultimately render comparison between or among countries less meaningful, these dimensions are nonetheless good starting points for analysis and further investigations for the Philippine AI scene.

Politically Japan's determination for self-sufficiency and to lead in practical applications of AI are seen as drivers of AI development. Japan's drive to automation and robotization is consistent with AI development. Its economic slowdown and talent shortage, however, are economic inhibitors of AI. Japan's elderly population provides an increased demand for AI in healthcare. The Japanese society is also seen as robot friendly. Japan's citizens take less issue with personal data processing for AI development purposes (Dirksen & Takahashi, 2020). In contrast, at least according to a representative of an AI startup who has done work in both Japan and the Philippines, the collection of personal data of Filipino citizens for machine learning is "harder". With respect to government-citizen relationship and a society's capacity to facilitate economic growth beyond blood ties, Japan is a high-trust society (Fukuyama, 1995), while the Philippines is not. Technologically, Japan's robotization of its industries takes a deeper hold in Japanese society and should be a boon to the overall AI development. But, just in most countries, cyber security threats continue to be a concern.

In the USA, to drive the advancement of AI are programs and measures enhancing "(1) healthcare and quality of life, (2) lifelong education and training, (3) business innovation and competitiveness, (4) accelerate[d] scientific discovery and technological innovation, (5) social justice and policy, and (6) national defense and security [transformation]" (Gil & Selman, 2019). Local parallel drivers *initially* identified in stakeholder consultations include:

(1) mandated use of electronic healthcare records (Appendix E),
(2) education for all AI users (Appendices D, E, & G),
(3) competitiveness of AI startups and businesses (Appendix G),
(4) enhanced research and development (R&D) programs, as supported by NEDA, NAST, and the academe (Appendices D & G),
(5) protecting human rights (Appendices D & G), and
(6) national security (Appendices D & G).

On the legal front, as AI development requires massive amounts of data, boon to such a development are Japan's support for Data Free Flow with Trust (DFFT) and its adoption of EU's Social Principles of Human-Centric AI in 2019 (Dirksen & Takahashi, 2020). In particular, a Japanese legal framework is being developed to facilitate the creation of data sharing platforms, especially for welfare and healthcare, allowing private and public sectors to exchange high-security database systems (Dirksen & Takahashi, 2020). Task 9 of the Philippine AI Roadmap aligns with this direction.

In our initial discussions with local stakeholders, the need to develop similar legal measures has been recognized. So is the need to put in "sunset clauses", especially in laws dealing with fast-moving technologies. Laws with specific technologies baked in, are vulnerable to technological obsolescence. These potential inhibitors of AI development must be attended to.

There is a general consensus among the local AI startups we consulted that Singapore's agnosticism towards algorithm, technology, sector, scale and business model in its AI governance framework is also good to maintain for the Philippines. While concerns might be raised with the Philippine AI Roadmap identifying agriculture, manufacturing, and services as priority areas for AI R & D (Department of Trade and Industry, n.d.), a charitable interpretation of these priorities is that of broad "starting points" rather than specific, exclusive directions undermining inclusivity. Nonetheless, AI startups perceive the general Philippine regulatory environment as being "too restrictive" already. If the AI framework being developed facilitates further restrictions, "it would only be a liability" (Appendix G). Exceptions to avowed "skepticism" towards specific technologies or sectors may have to be reexamined if having such technologies or sectors allow the country to "leapfrog" in becoming the industry leaders in those areas.

Some representatives from AI startups and the academe raised that dominance of foreign AI systems as a risk to national AI development, as allegedly it impedes local innovation. They tend to keep the Philippines as a mere training ground for foreign companies (Appendix G). With low AI competencies and lack of national industries, the Filipinos could remain mere buyers or users (rather than developers or producers) of AI technologies. The intensifying international transactions through AI places the country's national sovereignty at stake (Timmers, 2019). On a macro-level, AI can be another tool for colonization or control of the country, threatening national sovereignty, security, and privacy.

PCC representatives warn that this situation may also lead to unhealthy competition (Appendix F). This unhealthy competition extends to issues on various aspects of AI. Specifically, stakeholders raised that protection of their intellectual property rights is weak against companies with higher competition advantage (Appendix G). Indeed, the OECD concluded that AI competition may lead to collusion and even abusive conducts

(2021). These inhibitors must be identified and avoided through appropriate legislation and compliance. In this light, pro-active government regulations could play an important role in the development of AI, as they ensure compliance to international norms of competition and innovation. Competition policy-making entails "...ensuring regulatory frameworks support innovation and procompetitive AI applications without unnecessary burdens or competition barriers" (OECD, 2021).

In the end, different countries would measure differently to political, economic, societal, technological, and legal drivers and inhibitors of AI development. What may successfully drive one country to development will not necessarily do the same in another. In order to advance and support local AI development, stakeholders and the government have to follow through these drivers. Further studies are needed to validate these drivers and identify other potential drivers of AI development.

9. Recommendations

A national AI governance framework needs to be consistent with the national AI Roadmap already developed. But at the same time it should seek to influence the overall AI research and development in the country by providing the necessary ethical guardrails and governance principles for practices around the use of AI technologies. After a series of stakeholder consultations, key issues and concerns have been identified in at least these four (4) interrelated areas: digitization and infrastructure, workforce development, regulation, and research and development. Much of the strategic and operational details, however, remain to be worked out by concerned organizations and agencies, as an AI governance framework *qua* framework is limited to broader principles and elements.

9.1 Digitization and Infrastructure

(a) The framework should recognize that the national adoption of AI requires a robust connected and networked environment. A reliable and accessible internet connection is paramount to the successful development and adoption of AI in the Philippines. In line with this, Tasks 2 and 3 of the Philippine AI Roadmap, which are to be spearheaded by DICT and DTI, aim to improve internet quality and ensure that industries and support organizations have access to reliable and secure networks that are on par with acceptable global averages. A robust network environment is crucial to support, sustain, and scale up AI programs in the country.

(b) Generation and deployment of open data as fuel for AI development. Stakeholders are especially concerned about insufficient efforts to open public

data for the benefit of big data or AI research. Further, MSMEs and AI startups lament the power imbalance to larger foreign companies due to data ownership and exclusivity.

(c) In line with Tasks 7 and 8 of the Philippine AI Roadmap, the framework, through the leadership of DOST and DICT, should support the establishment of National Data Center (NDC) and National Research Cloud (NRC) to aid stakeholders, especially MSMEs, AI startups, and academic institutions, in data access and data value extraction. Data accessibility must be improved. Cross-sector data utilization should be supported, as it is the lifeline of AI.

9.2. Workforce Development

(a) Underline the imperative for the government to incentivize industries to offer Learning and Development (L&D) programs related to data extraction, data cleaning, data analysis, and machine learning, among others, for workers, managers, employers, and trainers. In this light, Task 23 of the AI Philippine Roadmap seeks to incentivize industries to send employees for graduate studies that focus on R&D to develop a scientific culture within organizations. Collaborative efforts from DTI, DOST, CHED, and DOLE are necessary to bring such plans to fruition.

(b) The framework should also emphasize the need to invest in AI-enabling resources and develop a deep appreciation for Science, Technology, Engineering, and Mathematics (STEM) and data science and analytics (DSA). In a parallel vein, Task 15 of the Philippine AI Roadmap aims to ensure the proper training of teachers in DSA and AI-centered graduate programs. This training is to be provided by DepEd, CHED, and DOST. Additionally, With Task 19, nurturing AI talents necessitates a collaboration with tech companies to provide sufficient resources and equipment to students and teachers alike.

(c) An AI governance framework needs to be built on top of DSA foundations (mathematics, statistics, and computing) in secondary education and general collegiate courses of data analytics, business analytics, and introductory AI in HEIs. Under the supervision of CHED, Task 18 of the Philippine AI Roadmap would also be about data visualization and storytelling as general education courses in universities.

(d) In pursuing AI, the framework should support the retooling or upskilling of government agencies anticipated to deal with AI governance and regulation. An efficient coordination of government efforts may be done through a body like the UK's Office for AI. While the development of AI requires a whole-of-society

approach, the challenges of streamlining and coordination attendant to such a massive mobilization of resources are too important to be left to potentially disparate, uncoordinated "organic" efforts by government agencies. A retooled or upskilled government is a sine qua non for AI to develop and truly be beneficial to the whole of society.

9.3. Regulation

(a) Adoption of inclusive growth, sustainable development and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; accountability, and trust as governance principles for AI in the Philippines.

(b) Endorsement of a Code of Ethics for the use of AI. Similar to Singapore's Implementation and Self Assessment Guide for Organisations (ISAGO), relevant areas to include are:
   (i) Regulatory risks
   (ii) Public relation risks (e.g. public perception towards the organization's AI practices)
   (iii) Deployment and adaptation costs
   (iv) Resources and internal champions to drive responsible implementation of AI

(c) The framework needs to highlight the development of a "conscientious" AI ecosystem. Intellectual property (IP) laws and data-protection laws have to be strengthened to address ethical concerns on the use of both data and AI. DICT, DOST, IPOPHIL, NPC, NAST, and CHR are the leads in this mission, which coincides with Task 27 of the Philippine AI Roadmap aiming to establish a committee of experts in data and AI ethics.

(d) Periodic revisiting and recalibration of the AI Roadmap and the subsequent governance framework. To be provided are details on timeline, mechanisms to achieve the 42 strategic tasks, groups or government agencies responsible, and budgetary requirements as well as "end of life" of specific tasks or versions. There appears to be low awareness of the Roadmap among stakeholders consulted. Better awareness and understanding of the Roadmap may lead to stakeholders having more to say about the framework being developed.

(e) Maintain a governance framework's general agnosticism towards specific algorithms, technologies, sectors, scale and business models, with special attention to the contextuality and time-sensitivity of government measures that aim to regular AI. Sunset clauses for specific regulations for emergent

technologies like AI are encouraged. Baking in specific technologies in legislation should be avoided.

(f) Endorsement of prohibited uses of AI, including: lethal autonomous weapon systems or killer robots, manipulative and exploitative practices, indiscriminate surveillance, and social scoring (Section 4.2). If AI legislation were imminent, according to stakeholders, a consensus around what AI should not be doing would not be hard to achieve.

(g) Governance and leadership by example. The government should have a programmatic deployment of AI in government agencies, programs, projects. Task 18 of the Philippine AI Roadmap states the need to "create quantifiable measures to track, coordinate, and improve government services and policies for industries." The imperative to power government with AI extends to the whole of society, not just industries.

9.4. Research and Development

(a) Align the framework with the forthcoming national innovation agenda. For instance, the framework's principles could help guide the country's innovation ecosystem and research practices towards the development of national innovation priority areas.

(b) Support the creation of the National Center for AI Research (NCAIR) that would serve as the country's hub for AI Research. As specified in the Philippine National AI Strategy Roadmap, NCAIR will be also responsible for advancing AI research and development, including algorithmic innovations, and nurturing AI talent and Data Science leaders. It will serve as the central hub for R&D collaboration between the academe, industries, multinational companies, and AI Startups.

(c) Broadened and deepened stakeholder consultations, especially to include sectors and groups most likely to be negatively impacted by AI. The use of diverse media and platforms to engage diverse stakeholders is also encouraged. Also encouraged is the representation in the company board of the sector most likely to be negatively affected by an AI system.

(d) The use and organization of AI governance elements (see Section 3 above) according to sectoral and organizational needs, allowing flexibility. Lobana's AI governance framework (2021) provides a comprehensive approach to the use of such elements.

(e) In pushing the boundaries of AI, the framework should incorporate Task 29 of the Philippine AI Roadmap that aims to improve the immediate recruitment of international talent and enhancement of international collaboration to increase international visibility. There is also a need to invest in AI R&D on strategic areas where the Philippines can perform well and compete globally. As required by Task 31 of the Philippine AI Roadmap and highlighted in the stakeholder consultations (See Appendix E), there should be strengthened academic-industry partnerships in AI R&D and technology transfer mechanisms.

BIBLIOGRAPHY

Alpaydin, E. (2021). Machine learning. Amazon. https://aws.amazon.com/machine-learning/what-is-ai/

Amnesty International. (2020, July 3). Dangerous anti-terror law in the Philippines yet another setback for human rights. Amnesty International. https://www.amnesty.org/en/latest/news/2020/07/philippines-dangerous-antiterror-law-yet-another-setback-for-human-rights/

Amazon Web Services. (2021). Correspondence, September 2021

Boucher, P. (2020). Artificial intelligence: How does it work, why does it matter, and what can we do about it? 1-60. doi: 10.2861/44572

Brutas, K. (2017). Social Media and Privacy: The Philippine Experience - Foundation for Media Alternatives. Foundation for Media Alternatives. https://fma.ph/resources/resources-gender-ict/social-media-and-privacy-the-philippine-experience/

Busuioc, M. (2020, August 15). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, 81(5), 825-836. DOI: 10.1111/puar.13293.

Castro, M.V. (2022) *AI Ethics and Governance Framework for the Philippines.* [Zoom presentation] Metro Manila, Philippines.

Cataleta, M. S., & Cataleta, A. (2020). Artificial Intelligence and Human Rights, an Unequal Struggle. CIFILE Journal of International Law, 1(2), 41–63. https://doi.org/10.30489/cifj.2020.223561.1015

China Aerospace Studies Institute. (2017). New Generation Artificial Intelligence Development Plan (DIGICHINA, Trans.).

Department of Trade and Industry. (n.d.). Artificial Intelligence Roadmap. http://innovate.dti.gov.ph/resources/roadmaps/artificial-intelligence/ (accessed 13 August 2021).

Department of Trade and Industry. (2021, 21 December). Stakeholders Consultation with AI Startups, Zoom Meeting

Department of Trade and Industry. (2022, 19 January). Stakeholders Consultation with Academe, Zoom Meeting

Deutsche Welle. (2021). Clearview AI controversy highlights rise of high-tech surveillance | DW | 14.06.2021. DW.COM. https://www.dw.com/en/clearview-ai-controversy-highlights-rise-of-high-tech-surveillance/a-57890435

Dietterich, T. G. (2017, October 2). Steps Toward Robust Artificial Intelligence. ResearchGate; Association for the Advancement of Artificial Intelligence. https://www.researchgate.net/publication/320185856_Steps_Toward_Robust_Artificial_Intelligence

Dirksen, N., & Takahashi, S. (2020). Artificial Intelligence in JAPAN 2020: Actors, Market, Opportunities and Digital Solutions in a Newly Transformed World. Netherlands Ministry of Economic Affairs and Climate Policy.

Doshi-Velez, F., Kortz, M., Bavitz, C., Gershman, S., O'Brien, D., Shieber, S., Waldo, J., Weinberger, D., & Wood, A. (2018). Accountability of AI Under the Law: The Role of Explanation. Harvard.edu. https://finale.seas.harvard.edu/publications/accountability-ai-under-law-role-explanation

Ernst & Young. (2021). Data foundations and AI adoption in the UK private and third sectors. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1010745/EY_DCMS_Data_foundations_and_AI_adoption_in_the_UK_private_and_third_sectors.pdf (accessed 31 December 2021)

European Commission. (2019). Communication From the Commission to the European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Building Trust in Human-Centric Artificial Intelligence.

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58496 (accessed 31 December 2021)

European Commission. (2020, Feb 19). White Paper on Artificial Intelligence – A European approach to excellence and trust. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed 31 December 2021)

European Commission. (2021). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Finch, L. (2020, July 8). UPR Info is making human rights recommendations more accessible—with a little help from machine learning - HURIDOCS. HURIDOCS. https://huridocs.org/2020/07/upr-info-is-making-human-rights-recommendations-more-accessible/

Floridi, L. (Ed.). (2021). Ethics, Governance, and Policies in Artificial Intelligence (Vol. 144). Springer International Publishing. https://doi.org/10.1007/978-3-030-81907-1

Foundation for Media Alternatives. (2016, March 6). State of Surveillance in the Philippines - Foundation for Media Alternatives. Foundation for Media Alternatives. https://fma.ph/2016/03/06/state-surveillance-philippines/

Fukuyama, F. (1995). Trust: The social virtues and the creation of prosperity. Free Press.

German Federal Ministry for Economic Affairs and Energy. (2018). Federal Government adopts Artificial Intelligence Strategy. https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20181116-federal-government-adopts-artificial-intelligence-strategy.html (accessed 18 December 2021)

Gil, Y., & Selman, B. (2019, August). A 20-Year Community Roadmap for Artificial Intelligence Research in the US. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI). https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/Community-Roadmap-for-AI-Research.pdf

Gutierrez, N. (2018, April 5). Did Cambridge Analytica use Filipinos' Facebook data to help Duterte win? RAPPLER. https://www.rappler.com/nation/199599-facebook-data-scandal-cambridge-analytica-help-duterte-win-philippine-elections/

Hall, H., & Khan, B. (2002). New Economy Handbook: Hall and Khan. NATIONAL BUREAU OF ECONOMIC RESEARCH.

Harrison, T., F. Luna-Reyes, L., Pardo, T., de Paula, N., Najafabadi, M., & Palmer, J. (2019). The Data Firehose and AI in Government. Proceedings of the 20th Annual International Conference on Digital Government Research. https://doi.org/10.1145/3325112.3325245

Human Rights Watch. (2021, September 15). Russia: Broad Facial Recognition Use Undermines Rights. Human Rights Watch. https://www.hrw.org/news/2021/09/15/russia-broad-facial-recognition-use-undermines-rights

Human Rights Watch. (2022, January 17). Philippines: End Deadly "Red-Tagging" of Activists. Human Rights Watch. https://www.hrw.org/news/2022/01/17/philippines-end-deadly-red-tagging-activists

IBON Foundation. (2016, June 29). Promote National Industrialization for National Development. IBON Foundation. https://www.ibon.org/promote-national-industrialization-for-national-development

International Committee of The Red Cross. (2019, August 20). Autonomy, artificial intelligence and robotics: Technical aspects of human control. International Committee of the Red Cross. https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control

International Service for Human Rights. (2020, July 21). Philippines | Anti-Terrorism Law further threatens the safety of human rights defenders. ISHR. https://ishr.ch/latest-updates/philippines-anti-terrorism-law-further-threatens-safety-human-rights-defenders/

Jacovi et al. (2021, March 03). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. DOI:10.1145/3442188.3445923

Japan Ministry of Economy, Trade and Industry. (2021, January 15). AI Governance in Japan Ver. 1.0: Interim Report. https://www.meti.go.jp/press/2020/01/20210115003/20210115003-3.pdf (accessed 16 Jan 2022)

Jose, D. (2022). Correspondence, 15 March 2022

Kertysova, K. (2018). Artificial Intelligence and Disinformation

How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered.
https://brill.com/view/journals/shrs/29/1-4/article-p55_55.xml?ebody=pdf-49903 (accessed 20 March 2022)

Li, H. (2020). Human Rights in the Age of Surveillance: China's Expansion of Technological and Normative Power.
https://as.nyu.edu/content/dam/nyu-as/ir/documents/Li-Huimin-ThesisFinal.pdf

Lobana, J. (2021). The Governance of AI-Based Information Technologies Within Corporate Environments (Unpublished doctoral dissertation). McMaster University, Canada.

Lobana, J. (2022). Correspondence, 27 January 2022

Lockey, S., Someh, I.A., & Gillespie, N. (2021, January). A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. DOI: 10.24251/HICSS.2021.664

Loi, M. & Spielkamp, M. (2021, May 05). Towards accountability in the use of Artificial Intelligence for Public Administrations. Algorithm Watch.
https://algorithmwatch.org/en/accountability-in-the-use-of-ai-for-public-administrations/

Maharjan, D.G. (2022) Remarks on the AI Governance Framework for the Philippines. [Zoom presentation] Metro Manila, Philippines.

Manyika, J., Silberg, J., & Presten, B. (2019, October 25). What Do We Do About the Biases in AI? Harvard Business Review.
https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

Miailhe, N. (2018). Competing in the Age of Artificial Intelligence: Current State of AI & Interpretation of Complex Data.
https://www.scor.com/sites/default/files/focus_scorartificial_intelligence.pdf.

Microsoft. Microsoft AI Principles. (n.d.)
https://www.microsoft.com/en-us/ai/responsible-ai (accessed 18 December 2021)

Millar, J., Barron, B., Hori, K., Finlay, R., Kotsuki, K., & Kerr, I. (2018, December 6). Accountability in AI: Promoting Greater Societal Trust.

Mint Press News. (2020, April 21). How the Us National Security State Is Using Coronavirus to Fulfill an Orwellian Vision. Mint Press News. https://mintpressnews.cn/national-security-state-using-coronavirus-push-artificial-intelligence-driven-mass-surveillance/266820/

Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education, Inc.

Occeñola, P. (2019, September 10). Exclusive: PH was Cambridge Analytica's "petri dish" – whistle-blower Christopher Wylie. RAPPLER. https://www.rappler.com/technology/social-media/239606-cambridge-analytica-philippines-online-propaganda-christopher-wylie/

OECD. (2019). Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText (accessed 27 Jan 2022)

OECD.AI. (2019). Korea, National Strategy for AI (2019) - OECD.AI. Oecd.ai. https://oecd.ai/en/wonk/documents/korea-national-strategy-for-ai-2019

OECD (2020), The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector, www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.htm

OECD. (2021). OECD Business and Finance Outlook 2021. OECD Business and Finance Outlook. https://doi.org/10.1787/ba682899-en

Olsen J. K. B., Pedersen S. A., and Hendricks V. F. (2009). A Companion to the Philosophy of Technology

O'Neill, O. (2002, April 24). Trust is the first casualty of the cult of transparency. https://www.telegraph.co.uk/comment/personal-view/3575750/Trust-is-the-first-casualty-of-the-cult-of-transparency.html (accessed 30 July 2020)

Open Data Philippines (ODPH). (n.d.) About Us. https://data.gov.ph/?q=about_us (accessed 13 August 2021)

Pacific Media Watch. (2018, March 22). How Philippine state surveillance is used as a tool to silence critics | Asia Pacific Report. https://asiapacificreport.nz/2018/03/22/how-philippine-state-surveillance-is-used-as-a-tool-to-silence-critics/

Personal Data Protection Commission Singapore. (2020). Model AI Governance Framework. 2nd edition. https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework (accessed 18 December 2021)

Philippine Statistical Authority. (n.d.). About OpenSTAT. https://openstat.psa.gov.ph/About (accessed 31 December 2021)

Philippine Innovation Act 2018 (Ph.).

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Levin, K. (2018). Artificial Intelligence & Human Rights: Opportunities & Risks. Harvard.edu. http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439

Salesforce. (2021). Asia Pacific AI Readiness Index 2021. https://www.salesforce.com/content/dam/web/en_au/www/documents/pdf/asia-pacific_ai-readiness-index-2021.pdf

Sanford, S. (2021, August 11). How to Build Accountability into Your AI. Harvard Business Review. https://hbr.org/2021/08/how-to-build-accountability-into-your-ai

Saslow, K., & Lorenz, P. (2019). Artificial Intelligence Needs Human Rights: How the Focus on Ethical AI Fails to Address Privacy, Discrimination and Other Concerns. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3589473

Shubhendu, S., & Vijay, J.F. (2013). Applicability of Artificial Intelligence in Different Fields of Life.

Shultz, G. (2020). Life and Learning after One Hundred Years; Trust Is the Coin of the Realm: Reflections on Trust and Effective Relationships across a New Hinge of History. Hoover Institute. https://www.hoover.org/sites/default/files/research/docs/shultz_finalfile_web-ready.pdf

Smith, C., & Brooks, D. J. (2013). Security Science. Elsevier Gezondheidszorg.

Smith, G., & Rustagi, I. (2021, March 31). When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Stanford Social Innovation Review. https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity (accessed 9 Jan 2022)

Steinhardt, J., & Toner, H. (2020, June 8). Why robustness is key to deploying AI. Brookings.
https://www.brookings.edu/techstream/why-robustness-is-key-to-deploying-ai/

Stoneman, P. & Battisti, G. (2010). Chapter 17 - The Diffusion of New Technology. *Handbook of the Economics of Innovation*, 733-760.
https://doi.org/10.1016/S0169-7218(10)02001-0.

Stop Killer Robots Campaign. (n.d.). Stop Killer Robots. Retrieved January 3, 2022, from https://www.stopkillerrobots.org/ (accessed 3 Jan 2022)

The White House Office of Science and Technology Policy. (2020). AMERICAN ARTIFICIAL INTELLIGENCE INITIATIVE: YEAR ONE ANNUAL REPORT.
https://www.nitrd.gov/nitrdgroups/images/c/c1/American-AI-Initiative-One-Year-Annual-Report.pdf

Timmers, P. (2019). Ethics of AI and Cybersecurity When Sovereignty is at Stake. Minds and Machines, 29(4), 635–645. https://doi.org/10.1007/s11023-019-09508-4

UK Center for Data Ethics and Innovation. (2020). AI Barometer Report.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894170/CDEI_AI_Barometer.pdf (accessed 18 January 2022)

UK Office for Artificial Intelligence. (2021). National AI Strategy.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf (accessed 28 December 2021)

United Nations. (n.d.). The Convention on Certain Conventional Weapons.
https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/ (accessed 4 Jan 2022)

U.S. Securities and Exchange Commission. (2017) Division of Investment Management. Guidance Update – Robo-Advisers.
https://www.sec.gov/investment/im-guidance-2017-02.pdf (accessed 18 Dec 2021)

Van Roy, V., Rossetti, F., Perset, K. and Galindo-Romero, L., AI Watch - National strategies on Artificial Intelligence: A European perspective, 2021 edition, EUR 30745 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-39081-7, doi:10.2760/069178, JRC122684.

Walsh, T. (2021, August 12). Lethal autonomous weapons and World War III: it's not too late to stop the rise of "killer robots."

https://theconversation.com/lethal-autonomous-weapons-and-world-war-iii-its-not-too-late-to-stop-the-rise-of-killer-robots-165822 (accessed 4 Jan 2022)

World Economic Forum. (2021). Human-Centred Artificial Intelligence for Human Resources: A Toolkit for Human Resources Professionals. https://www3.weforum.org/docs/WEF_Human_Centred_Artificial_Intelligence_for_Human_Resources_2021.pdf (accessed 31 December 2021)

Whittaker, M. (2020, November 2). Who am I to decide when algorithms should make important decisions? The Boston Globe. https://www.bostonglobe.com/2020/11/02/opinion/who-am-i-decide-when-algorithms-should-make-important-decisions/ (accessed 7 Jan 2022)

Young, E., Wajcman, J., & Sprejer, L. (2021). Where are the Women? Mapping the Gender Job Gap in AI. Policy Briefing: Full Report. The Alan Turing Institute, 2021–03.

Zewe, A. (2022, February). Can machine-learning models overcome biased datasets? MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2022/machine-learning-biased-data-0221

Appendix A

AI Governance Principles in Selected Countries, as Reflected in their AI Frameworks, Strategies and Plans, Regulatory Proposals

Key: ☒Not mentioned, ☑Mentioned only, 🔷 Covered

	Singapore[2]	China[3]	Japan[4]	Korea[5]	UK[6]	EU[7]	US[8]
Inclusive growth, sustainable development and well-being	☒	☑ "social development" "struggle toward the goal of a moderately prosperous society"	☒	☑ "establishing an inclusive job safety network" "creation of sustainable AI ecosystem"	☑ "health and social care…", "improve lives"	🔷 "Ensuring inclusiveness in the AI-driven economy" "sustainable development" "safeguard inclusion and social welfare"	☑ "Economic growth" "development for the good of American people…enriching lives"
Human-centered values and fairness	☑ "appropriate degree of human involvement"	☑ "implement people-centric development thinking"	☒	☑ "nurturing human talents"	☑ "involvement of diverse talents and views of society" to achieve goals "	🔷 "Human capital" "human rights" "fair data governance and use" "fair competition	☑ "human-AI collaboration"
Transparency and explainability	☑ "open and transparent communication"🟦	☑ "Open and transparent AI supervision system"	☒	☒	☑ "trustworthiness, adaptability, and transparency"	🔷 "Legislative transparency and governance" "explainability [for trustworthy AI]"	☑ "Priority areas of AI R&D emphasize the development of explainability mechanisms that help human users understand reasons for AI outputs, along with methods to test, evaluate,

[2] Personal Data Protection Commission Singapore, 2020. Model AI Governance Framework. 2nd edition.
[3] China Aerospace Studies Institute, 2017. New Generation Artificial Intelligence Development Plan (DIGICHINA, Trans.)
[4] Japan Ministry of Economy, Trade and Industry, 2021. AI Governance in Japan Ver. 1.0: Interim Report.
[5] OECD.AI. (2019). Korea, National Strategy for AI (2019) - OECD.AI. Oecd.ai. https://oecd.ai/en/wonk/documents/korea-national-strategy-for-ai-2019
[6] UK Office for Artificial Intelligence. (2021). National AI Strategy.
[7] Van Roy, V., Rossetti, F., Perset, K. and Galindo-Romero, L., AI Watch - National strategies on Artificial Intelligence: A European perspective, 2021 edition, EUR 30745 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-39081-7, doi:10.2760/069178, JRC122684.
[8]The White House Office of Science and Technology Policy. (2020). AMERICAN ARTIFICIAL INTELLIGENCE INITIATIVE: YEAR ONE ANNUAL REPORT. https://www.nitrd.gov/nitrdgroups/images/c/c1/American-AI-Initiative-One-Year-Annual-Report.pdf

							verify, and validate their performance.”
Robustness, security and safety	✅ ”explainability, robustness, regular tuning”	✅ “elevate national defense strength and assure and protect national security”	☒	✅ “create the safest AI-use environment in preparation for dysfunction and security threats that may occur due to spread of AI” “strengthening public safety”	✅ ”delivering on safety, security and trust.”	🟦✅ “robust digital ecosystem for AI” “safe and secure cyberspace”	✅ “National security and defense” “safety critical systems” “improve worker safety”
Accountability	✅ ”building good accountability practices”	✅ “establish traceability and accountability system”	☒	☒	☒	✅ “Accountability of AI actors”	☒
Trust	✅ ”build a trusted and progressive AI environment”	✅ ”promote social interaction and mutual trust”	✅ ”reinforce mutual trust among companies, users and the government in society.”	✅ ” national policies for trustworthy AI”	🔷 ”establishes the most trusted and pro-innovation system for AI governance in the world”	🔷 “Trustworthy AI deployment”	🔷 “To garner trust and confidence, AI technologies should be transparent in how they work and provide reasonable guarantees on the safety, security, robustness, and resiliency of their operation” “building public trust and confidence in AI technologies”

Appendix B

Developing a National AI Framework: Consultation/Discussion Questions for AI business startups,[9] the academe, government, business.

1. Does your organization subscribe to fairness; accountability and auditability; reliability, safety, and security; transparency as the principles guiding your AI systems or AI-augmented practices? (See ai-ph.org/devframe for details on these principles.) Are they the appropriate values to inform our own national AI (artificial intelligence) governance framework? Do such principles add value to your products or services, academic programs? Please explain.

2. Pick three (3) elements of AI governance (see ai-ph.org/devframe for the elements and their descriptions) you find most relevant to your organization. Please indicate your answers using this form.

3. In your view, which current local or national laws and regulations (e.g., competition, data privacy, consumer protection, intellectual property) tend to facilitate or impede the development of your AI systems or AI-augmented practices? What specific laws or regulations do you want to see enacted? For instance, do holders of big data used for professional AI practice need more concrete protection against theft and unauthorized use or divulgence of such data?

4. In your professional opinion, what are the proper roles of specific government agencies, industries, academia, or other stakeholders in the development of AI?

5. The Singapore AI governance framework maintains an agnosticism towards algorithm, technology,[10] sector, scale and business model. (See "Preamble" of the SG AI framework.) Do you see this as an appropriate stance for our own AI framework?

---

[9] Version 1 (21 Dec 2021, for AI Business Startups; 19 Jan 2020, the academe). Shortcut to its current page: ai-ph.org/discuss. All versions and the responses they have elicited in consultation meetings have been archived accordingly.

[10] An example of a specific technology being baked into law is referenced in RA 9239. Originally intended "to ensure the protection and promotion of intellectual property rights", the law (via its IRR) requires any person, establishment or entity "offering to the public with intent to profit the use of optical disc writers and rewrites," to register with the Optical Media Board (OMB). The unintended, unanticipated consequence of this regulation is the inability of Filipino doctors, hospitals or clinics to *legally* share diagnostic images with their patients. Not only are these entities mostly unaware of such a requirement, it is also cumbersome (if not altogether impractical) to register machines capable of writing on optical disks. When the law was crafted, optical disks were the state of the art for storage media for sharing.

6. **Risk** = Magnitude (Severity) of Harm x Probability (Likelihood) of Harm. Severity and likelihood range from 'negligible' (value=1), 'limited' (2), 'significant' (3), and 'maximum' (4). Please identify top (3) risks in **your domain** of AI application, indicating their respective scores. Example: application: staff recruitment; risk: data privacy breach = 6 (magnitude=3, likelihood=2). For "common risks", see also Section 7 of ai-ph.org/devframe.

7. The European Commission proposes to prohibit the use of AI for (a) manipulative and exploitative practices, (b) indiscriminate surveillance, and (c) social scoring. Our local AI business startups added lethal autonomous weapons systems (LAWS) to be prohibited (Sec 4.2 of ai-ph.org/devframe). Are these prohibitions to be stipulated in our national AI framework justified? Why or why not?

Appendix C

Participants in Stakeholder Consultations

Dec 21, 2021; AI Startups: Roby Alampay (Pumapodcast), Imelda Andales (Symphonics (Tacloban)), Roger Collantes (Asian Institute of Digital Transformation), John Duenas (Hybrain), Richard Ebdalin (Dirtbag (CDO)), Tony Intal (DK PO Fulfillment Company Inc.), Dustin Masancay (Third Derivative), Spark Perreras (Pearlpay)

Jan 19, 2022; Academe: Dr. Chris Jordan Aliac (Cebu Institute of Technology University), Dr. Rowell Guillermo Atienza (University of the Philippines Diliman), Dr. Jocelyn B. Barbosa (University of Science and Technology of Southern Philippines), Dr. Randy S. Gamboa (University of South-Eastern Philippines), Dr. Yvette G. Gonzales (Iloilo Science and Technology University), Dr. Eugene Rex L. Jalao (University of the Philippines Diliman), Dr. Erika Fille T. Legara (Asian Institute of Management), Dr. Prospero C. Naval Jr. (University of the Philippines Diliman), Dr. Karl Erza S. Pilario (University of the Philippines Diliman), Dr. Miguel Francisco M. Remolona (University of the Philippines Diliman)

Feb 3, 2022; Government Agencies: Acd. Rhodora V. Azanza (NAST), Katrina M. Atienza (NEDA), Usec. Benjo Santos M. Benavidez (DOLE), Dir. Angelo M. Benedictos (BITR), Dir. Vina Liza Ruth C. Cabrera (IPOPHL), D.D. Aldryn Consolacion (BSP), Kevin S.M. Cuevas (NEDA), Dir. Adeline de Castro (DOLE), Demark Descatiar (DOLE), Corinne Escartin (CHR), D.D. Justin Ray Angelo Fernandez (BSP), A.G. Lilia C. Guillermo (BSP - TDIO), Dir. Noel Guinto (BSP), Jonathan Juan (CHR), D.D. Agnes Perpetua R. Legaspi (BSP - TDIO), Graciela E. Mante (NEDA), Cherrie R. Mapa (BSP), Dir. Charles Merioles (IPOPHL), Allan D. Mordeno (PCC), May-Rose Pariñas (DOST), Dir. Gemma Parojinog (CHR), Jonathan Y. Ragsag (NPC), Clarinda G. Reyes (DOST), Mariane Reyes (NEDA), Rossvern Reyes (BSP), Hazel Sambo (NEDA), E.D. Kenneth V. Tanate (PCC), Ivy Grace T. Villasoto (NPC), Dir. Archellis A. Villena (BSP - TDIO)

Feb 24, 2022; Businesses: Christine Bata (Global In-House Center Council), Jomar Bernedo (Philippine Software Industry Association), Melvin Jeffrey Chan (PLDT Enterprise), Jay Chavez (Ionics EMS), Cristina G. Coronel (Healthcare Information Management Association of the Philippines), Laurence Cua (SM Investments Corporation), Lorenzo Fabic (Global In-House Center Council), Antonio Geniston (AG Pacific Neutriceuticals), Celeste Ilagan (IT & Business Process Association of the Philippines), Dominic Ligot (Analytics Association of the Philippines), Natalie Lim

(Global In-House Center Council), Jack Madrid (IT & Business Process Association of the Philippines), Jeffrey G. Mijares (Samahan sa Pilipinas ng Industriyang Kimika), Eduardo G. Mulong (Management Association of the Philippines), Sherwin M. Pelayo (Analytics Association of the Philippines), Rex Victor Puentespina (Malagos Agri-Ventures Corporation), Jay Santisteban (Contact Center Association of the Philippines), Antonio Ll. Sayo (Employers' Confederation of the Philippines), Ben Teehankee (Analytics Association of the Philippines), Dr. Francis Aldrine A. Uy (USHER Technologies), Tony Villaflor (Semiconductor and Electronics Industries in the Philippines, Inc), Lito Villanueva (Rizal Commercial Banking Corporation)

March 10, 2022; Civil Society Organizations: Dr. Jose Ramon G. Albert (Federation of Free Workers), Mr. Roberto Bacson (Institute of Corporate Directors), Ms. Lindsay A. Barrientos (UNESCO Philippine National Commission), Atty. Paula Sophia G. Estrella (Ateneo Human Rights Center), Ms. Lisa Garcia (Foundation for Media Alternatives), Mr. Carlos Jose P. Gatmaitan (Institute of Corporate Directors), Mr. Reynaldo Antonio D. Laguda (Philippine Business for Social Progress), Mr. Eric Rosales (Institute of Corporate Directors), Dr. Charlotte Justine Diokno-Sicat (Philippine Economics Society), Ms. Carla Angeli A. Ronquillo-Solis (Institute of Corporate Directors), Mr. Rex A. Ubac Jr. (UNESCO Philippine National Commission)

March 17, 2022; Microsoft: Dale Jose, Joanna Velez Rodriguez, Martie Valmores

May 17, 2022; Government Agencies: USec. Maria Victoria Castro (Department of Information and Communications Technology), Director Diane Gail Maharjan (National Economic Development Authority), Ms. Clarinda Reyes (Department of Science and Technology (Philippine Council for Industry, Energy and Emerging Technology Research and Development)), Ms. Mary Rose Parinas (Department of Science and Technology (PCIEERD), Mr. Ramon Abraham Sarmiento (Bangko Sentral ng Pilipinas), Mr. Rossvern Reyes (Bangko Sentral ng Pilipinas), Ms. Grace Del Rosario (Bangko Sentral ng Pilipinas), Director Lizzie Cabrera (Intellectual Property of the Philippines), Assistant Director Charles Merioles (Intellecctual Property of the Philippines), Atty. Ivy Grace T. Villasoto (National Privacy Commission), Mr. Jonathan Rudolph Y. Ragsag (National Privacy Commission), Mr. Mark Jeson L. Pura (National Privacy Commission), Mr. Vincent Leo L. Gamboa (Commission on Human Rights), Ms. Rica Beatrice Ambubuyog (Commission on Human Rights). Academe: Mr. Rowell Guillermo Ma. Atienza (University of the Philippines Diliman), Mr. Rafael A. Cabredo (De La Salle University), Ms. Yvette G. Gonzales (Iloilo Science and Technology University), Dr. Karl Erza S. Pilario (University of the Philippines Diliman), Mr. Miguel Francisco M. Remolona (University of the Philippines Diliman). Industry: Dr. Adrienne Heinrich (UNIONBANK), Mr. Jack Madrid (IT & Business Process Association of the Philippines), Ms. Michelle Alarcon (Analytics Association of the Philippines), Mr. Wilfredo "Dodong" Montino, Jr. (UNIONBANK), Mr. Sherwin M. Pelayo (Analytics Association of the Philippines), Mr. Monchito Ibrahim (Analytics Association of the Philippines), Mr.

Dominic Ligot (Analytics Association of the Philippines), Mr. Alexis Fante (Ayala Health), Mr. James Arnold Faeldon (Ayala Corporation), Ms. Jiannah Beatrice Rosal (Employers' Confederation of the Philippines), Mr. Robert Maronilla (Employers' Confederation of the Philippines), Ms. Celeste Ilagan (IT & Business Process Association of the Philippines), Mr. Frankie Antolin (IT & Business Process Association of the Philippines), Mr. Coco Alcuaz (Makati Business Club), Mr. Lito Villanueva (Rizal Commercial Banking Corporation), Mr. Robert Alexander Campos (Rizal Commercial Banking Corporation), Mr. Jennie Reifsnyder (Rizal Commercial Banking Corporation), Dr. Danilo Lachica (Semiconductor and Electronics Industries in the Philippines, Inc.), Ms. Clarisse Maludon (Semiconductor and Electronics Industries in the Philippines, Inc.), Ms. Rose Duazo (Semiconductor and Electronics Industries in the Philippines, Inc.), Mr. Rhett Ramos (Allegro MicroSystems), Mr. Wowie Makalintal (Gruppo EMS), Mr. Jonathan Ac-ac (Amkor Technology Phils. Inc.), Mr. Jonathan Mondero (Amkor Technology Phils. Inc.), Mr. Harold Carlo Rebuldela (Onsemi), Ms. Norika Pineda (US ASEAN Business Council Inc.), Mr. Jason Albia (Intervenn Biosciences). Startups: Ms. Cherry Murillon (cawil.ai), Mr. Tony Intal (DK PO Fulfillment Company Inc.), Mr. Spark Perreras (Pearlpay), Mr. Allan Tan (Monstar Lab), Mr. Gian Paulo dela Rama (aiah), Ms. Kathleen Yu (Rumarocket), Ms. Patricia Opleda (Senti AI), Mr. Carlo Delantar (Core Capital). Civil Society Organizations: Ms. Lindsay A. Barrientos (UNESCO Philippine National Commission). Big Tech: Mr. Dale Pascual Jose (Microsoft Philippines), Ms. Annabel Lee (Amazon Web Services - Philippines).

Appendix D

Concerns from Stakeholders on AI Principles

	AI Startups	**Academe**	**Government Agencies**	**Businesses & Big Tech***	**Civil Society Organizations**
Inclusive growth, sustainable development and well-being	Clear governance achieves compliance without operational costs	AI's increased capability for synthetic messaging impacted the BPO industry	AI enables national development (NAST) The National Innovation and Strategy Document involves AI (NEDA) "AI has to be inclusive in order to speed up the fulfillment of AmBisyon 2040" AI needs to be consistent with the PH Competition law and applicable to all industries to ensure fair competition There should be no barriers in entering the market (PCC) AI could diminish the digital divide and increase digital access, leveling the playing field to some extent.	Core values such as shared prosperity and sustainable value creation lead to sustainable growth and development. AI could usher local growth and enable global competitiveness among our industries. AI models, which impact jobs and employment, should be regulated to provide social protection to workers. Adopting AI in local industries can produce more jobs and drive investments An AI strategy to adopting innovative technologies can be encouraging organizations and citizens through incentivizes (Amazon)	Regulatory frameworks, compliance, and accountability mechanisms ensure inclusivity (PIDS) Make AI development a job-creating system for workers (FFW) Learner-driven approaches and other growth strategies

				AI should benefit all (Amazon) Democratize AI technology to include PWDs and other marginalized sectors (Microsoft) Cloud First Policy (Microsoft)	
Human-centered values and fairness	There is a need for clear definition on fairness to ensure compliance It would be helpful to define fairness from a technological perspective. Ambiguous definition of fairness could decrease value for start-ups.	Data and algorithms have their own predisposition which must be assessed to guarantee fairness There should be commitment in ensuring an unbiased data set used in training AI models . On top of ISO standards, it should also be guaranteed that AI tools are unbiased, accurate, and ethical. Fairness is not an all encompassing term; it is recommended to identify a term that includes inclusivity, equality, etc.	AI has an impact on political rights where particular sectors are vulnerable to manipulators; thus, regulations must be for protection of rights (CHR)	AI should be beneficial to the people and not just a selected few. The power of AI should be harnessed to address societal problems and advance human welfare and quality of life. There should be more stringent laws on the use of AI-models that dictate what is fair and what is not Address workforce upskilling issues (Amazon) Responsible AI (Microsoft)	Stakeholder-driven approach where each sector has the chance to participate in strengthening AI (FMA) Human rights perspective and frameworks must inform AI gov't framework (AHRC)
	Emphasized the necessity of human involvement, especially in the risk of displacement of humans in the labor force.				
Transparency and explainability	Explainability is necessary, which entails clear communication of AI to stakeholders Transparency entails education for accurate	AI as a blackbox system contradicts with transparency Transparency must have its own qualifiers (PNaval)	Issues in private intervention on data, data integrity, and discomfort in blackbox approaches where stakeholders are concerned with replicability of results		Need for information entails collection of good quality data (PES) Create platforms to ensure all sectors (eg persons with disabilities,

	interpretation and AI utilization Not all decisions or not all AI elements should require explainability or transparency AI companies should be able to explain how the models are built in order to provide the client with crucial information on how to use AI and how the AI comes to its conclusions. Transparency revolving around human-centricity is a critical aspect. Explainability should have different levels, given the varying standards each industry has, to ensure that other AI players would not be crippled. It wouldn't make sense to consider explainability for AI dealing with less risky or trivial tasks. Explainability is necessary to avoid instances where the end-user creates their own interpretation because of not being able to understand the implications of the AI's conclusions. The deeper the neural network, the more difficult	Explicability must be added to clearly define transparency. If AI is not comprehensible by experts, given that they are blackbox systems, then it would be more obscure for third-parties. Transparency could not be included because this principle contradicts with the principles of blackbox models set forth by DTI (PNaval). Transparency is only necessary for certain elements of AI such as policies and processes, but not the data.	need to be addressed (BSP)		underprivileged, etc.) can understand AI (FMA)

	explainability would be even for domain experts.				
Robustness, security and safety	Security should also be a priority since breach of data is catastrophic.	Security should go down to the level of the algorithm being used by the AI.	Ensure the principle of robustness, security, and safety of AI is rooted in regulations (NPC). Data masking tools are established to prevent leakage of private information (BSP)	Rainforesting, implement a Cloud First Policy (Microsoft) Autonomous AI systems should have a 'kill-switch' or safeguards to ensure ultimate controllability by humans (Microsoft)	
Accountability	Data Privacy contradicts with auditability There should be clear-cut guidelines with regards to who would review and the scope of the review Incorporate a Human-in-the-Loop process (HTL) such that there's always a way to check what the results of the AI mean and its implications.	AI as a blackbox system cannot be audited (PNaval). Applications produced must be contextualized. There should be encompassing laws on a national level that would determine who's accountable for the consequences of AI. The lack of understanding on the inner workings of AI blackbox systems may impede in adhering to the principle of accountability and auditability.		From a regulatory standpoint, there is no way of assigning liabilities for machines that have agency. Current laws are static in nature, sort of binary, which do not capture the nature of AI being all probabilistic	Institution governing regulatory frameworks (PES) Standards ensure harmony of local laws with international laws (FMA)
Trust		Human-intervention on AI Framework is important. Human-in-the-loop process (HTL) should be present.	Expectation of PCC for AI functions to be trustworthy (PCC), which means it acts in accordance with existing AI frameworks.	AI could make privacy obsolete. Trust First, Technology Second (Microsoft) Dialogue with citizens is important to foster trust in AI and gain their buy-in	Issues on trust stem from unawareness or lack of information

			Build trust and demystify AI through education There should be guidelines established regarding the collection of personal data (NPC) AI should adhere to ethical requirements and human values wherein humans are the one in control of AI (CHR).	over time (Microsoft). AI systems should perform reliably and safely to foster trust	

*multiple sessions held with businesses and big tech representatives. Position papers from Amazon and Microsoft have also been submitted for consideration.

Appendix E

Concerns from Stakeholders on Governance Structures

	AI Startups	**Academe**	**Government Agencies**	**Businesses & Big Tech**	**Civil Society Organizations**
AI Startups	Partner programs for AI companies and start-ups would be beneficial. Centralized data management for Philippine AI companies to learn from each other would be helpful in scaling training and internal transfer learning.	AI Startups must have their own department for research and development In charge of innovation and expansion. Needs funding. Lead in AI utilization.		Develop AI models for all sorts of application that is useful in regions with varying markets Champion the country as an investment for our principal	
Academe	Research is too academic; there is a need for more research geared towards industries Academe holds a key role in maintaining balance between research and commercialization. Academe should be there to provide the skills and training to AI graduates.	More research-based AI training and development programs Produce AI graduates based on standards set by government agencies	Education on AI is vital for the protection of human rights. It would be beneficial to build more educational programs in AI to help raise awareness.	Have anticipatory moves when it comes to the R&D of new technologies. Continuously update the curriculum of AI-related courses in higher studies and technical schools. Research in universities regarding the various stages of AI development is necessary. Scholarship programs on all sorts of digital transformation courses including AI	Comprehensive accounting of AI (ICD) Interpretation and communication of data (ICD)

				Advanced knowledge and skills in all levels in society Build relevant curriculum on the required competencies on AI	
Government Agencies	Train labor force In charge of accreditation for AI workforce and practitioners to ensure that standards are met. Develop ways to centralize government data to provide a common source of truth for training AI models. A whole government approach is necessary in order for the AI industry to advance. DOST should spearhead AI-research or AI related innovations . DTI should handle the business side of AI development and adoption with special focus on SMEs.	In charge of funding In charge of legal and policy frameworks to create a healthy local ecosystem for AI initiatives Serve as a liaison between enterprises and academia in various DSAI initiatives Provide right environment and infrastructure Set the minimum standards for AI workforce Spearhead certification programs necessary for the vetting and assessment of authenticity of AI workforce	National Innovation Council (NIC) through NEDA is in charge of Coordinated and Harmonized Policy Advisory NPC must be involved in various standardization efforts globally CHR oversees auditability of AI Systems for humanitarian work and checks if there are undue denial of rights due to the use of AI. Philippine Statistic Authority and NPC could aid AI Startups in providing sufficient data for AI models. Improve the quality and soundness of data in government archives for AI models to learn effectively.	Adopt more enabling laws that would not stifle AI development and innovation Have certification programs for AI practitioners and AI companies DOH and/or PhilHealth need to mandate the use of EMRs. Promote and develop data infrastructure (hardware) to expedite digital transformation amongst PH companies Reduce the need for paperforms in order to capture more electronic data. Invest heavily in AI technology to attract more foreign investments Invest in infrastructure to support AI technology. Public sector as first mover to drive trust in AI and public adoption	Less governance elements, easier to monitor (ICD) Minimal approaches to the elements eg. COBIT, ISO, etc. (ICD) Create baseline indicators to assess impact of elements (PES) NPC in charge of implementation (AHRC) Need to be careful in creating more offices since current offices are not working (PIDS) Ensure presence of regulatory framework and institutions (PES) Work within the laws; assessment of existing agencies will help in managing AI governance (PES) NPC in charge of data sharing and handling (ICD)

				(Amazon) AI does not have to be government-led, it can be government-supported (Microsoft) Should take the lead in the implementation of Cloud First policy (Microsoft) Gov't should be flexible on the budgeting for AI technology.	

Appendix F

Concerns from Stakeholders on Risk Management

AI Startups	Academe	Government Agencies	Businesses & Big Tech	Civil Society Organizations
Adoption of data privacy Bias Lack of sharing of data and best practices Security breaches	Disinformation Misinformation Loss of trust in AI Undervaluation of public data Improper use of AI models Unqualified AI graduates from HEIs	Vertical ownership creates unhealthy market competition (PCC) Indiscriminate surveillance against government critics used to harass them online (CHR) Profiling a risk for national security (CHR) Discrimination and disinformation (CHR) Anti-competitive practices involve surveillance to gain more profit (PCC)	Data breach and hacking Leak of off-grid data and information AI could make privacy obsolete Psychological Manipulation (MAP) Exploitation Misalignment between knowledge and management could render AI technology disruptive Autonomous AI could destroy numerous existing businesses or industries Identity theft Unwanted targeting of AI (the ability to identify someone even without explicit identifiers and consent) False dilemmas	Lack of data sharing facilitates lack of understanding (PIDS) Use of privacy for control rather protection (PIDS)

			Disparate treatment and impacts of AI models Abuse and unethical use of AI Disinformation in financial and healthcare services AI-propagated rumors can trigger a bank run, or a market bubble and crash AI regulation should be risk based and liability clearly assigned to actors to address risks (Amazon) Cybersecurity Only through adoption of AI can risks be discovered (Microsoft) Biases in data used to train AI algorithms may perpetuate inequality and discrimination (Microsoft)	

Appendix G

Concerns from Stakeholders on Inhibitors and Drivers of AI Development

	AI Startups	Academe	Government Agencies	Businesses & Big Tech	Civil Society Organizations
Inhibitors	The Philippines as mere training ground for foreign companies is a risk to the country's AI sustainable development Data privacy and consumer protection regulations make it difficult to develop AI systems, since compliance makes it harder to collect data sets and train software. Nevertheless, regulations push the industry to act more carefully and ethically. Legislation is too inflexible for AI which is constantly evolving. Restricting the Philippines with specificities is a liability.	Lack of local AI competencies Substantial foreign ownership of AI services is a risk of "colonization" through AI. This is also a threat to national sovereignty, security, and privacy. The Filipnos being mere buyers or consumers of AI products and services rather than AI developers themselves. Lack of appropriate national laws is a threat to accountability and auditability Absence of a national body dedicated for AI leaves various AI stakeholders without guidance Existing policies on various AI systems differ from one another Difficulty to protect the academe's intellectual	Confusion on the scope of Data Privacy Act of 2012 (NPC). Singaporean agnosticism may contribute to the widening social inequality (BSP).	Most businesses claim there are no legal conflicts in the development of AI. Hacking is seen as a threat. Current issues in cybercrime and even internet bandwidth must be addressed in preparation for AI development and adoption. Various types of data must be identified. Some should not be regulated; overregulation may impede AI development. Guidelines and regulations must be balanced. Current IT governance frameworks do not capture systems that could dynamically adjust and learn on their own. AI could accelerate moral	No need to add more laws (AHRC) Lack of understanding data science may render tools useless (ICD) Difficulty in digitizing data through data gathering (ICD) Difficult to effectively implement prohibition of AI uses through law (UNESCO)

		property rights against companies that have greater financial capacity. This leaves our researchers vulnerable to intellectual property rights violations. Current data availability in the Philippines is extremely inadequate Tedious data processes hinder ease of obtaining, analyzing data. Commercialization of data must also be discussed. Data Privacy Law limits access to useful data in organizations. Access to data is too restrictive, exceptions to the Data Privacy Law are not implemented well (Rex)		hazards which may encourage monopoly of industries. The lack of legal liability framework is an inhibitor of AI development. Current laws do not capture the nature of AI as probabilistic AI may contribute to making privacy obsolete since current data privacy laws do not account for AI identifying individuals. If AI knowledge is too specialized and the management is not as technically informed, problems may occur. The scope of control by government should not be implemented too early as it might impede AI innovation.	
Drivers	Being seen as having adequate GDPR levels of data protection to process EU data. In policy making, sunset clauses are helpful for fast-evolving AI. Innovation is prioritized.	Legislation ensures protection of people, data and information, algorithms, technology, and other components of AI. Legislation must clarify the rights of each stakeholder from data owners to data holders. Data-availability is championed on the premise that there is still ownership	Lawful interventions must be critically mandated because it can facilitate or impede AI development (IPOPHL). AI and Machine Learning Guidelines and Standards aid AI development	Laws must be flexible and inclusive for faster adaptation of local emerging technologies. Enactment of certain laws will accelerate the development of AI. Data ownership must be protected; algorithms or models come secondary	HR and AI auditing (AHRC) Strengthening internet connection (ICD)

		A law to have organizations provide non-identifying personal data that can help government make better decisions and policies through research institutions (Rex)	Social license in terms of responsible AI design and implementation helps	relative to data. Large-scale data that are stored off-grid ensure protection of data privacy and full control over data processes. Laws must have corresponding penalties that would discourage and prevent instances of hacking. Education allows clients and end-users to know that they are protected and their data is secured. Data availability makes AI models more accurate. AI must develop all sorts of applications that could be applicable and useful in regions with varying markets. Competitiveness and benchmarking must be included in the legal policy framework. Consumer protection laws should include protection against psychological manipulation and mental health issues and must also provide specific provision for the youth. Enable cloud-first and	

				pro-innovation policies (Amazon) Research and development (Amazon) Democratizing data, technology, AI (Microsoft) Promote AI-ready culture (Microsoft) Ensure that the framework accelerates AI through cloud services (Microsoft) Digital sovereignty (Microsoft) Continuous dialogues with citizens to ensure AI is suitable for the Philippines	

Appendix H

Glossary of Terms

Artificial Intelligence (AI) refers to a range of technologies and systems performing tasks and solving problems commonly associated with human intelligence (Alpaydin, 2021).

AI-Solutions Providers are individuals or groups that develop and provide solutions, systems or applications using artificial intelligence (Personal Data Protection Commission Singapore, 2020).

Audit is a process carried out by an independent body that involves the probing, understanding, and reviewing the behavior of algorithms (Personal Data Protection Commission Singapore, 2020). A meaningful auditing process typically has three (3) phases: information, explanation or justification, and (a possibility of) sanctions (Busuioc, 2020).

Autonomy is the capacity of the user to exercise self-governance or act in accordance to one's own capacity for decision-making.

Algorithms refer to finite sequences of rigorous instructions used as specifications for performing calculations and processing data. It may also refer to precise rule-based procedures that a computer could follow, step-by-step, to decide how to respond intelligently to a given situation. Explainable algorithms allow for higher levels of confidence in AI, as its end-users gain better understanding of how it makes decisions in the face of wide-ranging and uncertain variables that interact with each other (Boucher, 2020).

Artificial Neural networks are computing systems composed of layered and interconnected clusters of nodes. Inspired by the biological neural networks that model brains, these networks "learn" to perform tasks by considering or abstracting from examples or instances, commonly without being programmed with any task-specific rules (OECD, 2020).

Diffusion is the process by which something new (say, technology) spreads throughout society (Hall & Khan, 2002). It may also be viewed as the phenomenon by which the market for a new technology changes over time and from which production and usage patterns of new products and production processes result (Stoneman & Batisti, 2010).

Governance framework provides a mechanism for organizations and AI-solution providers to have a clear understanding and oversight of one's expectations, objectives, risk appetite, accountabilities, and responsibilities (Smith & Brooks , 2013).

Machine Learning is one of the main ways AI is being developed and applied, with algorithms that learn and adapt through increasing amounts of training data which correspondingly improves their performance over time. It involves decision trees and Bayesian networks, thereby allowing for more transparency in its decision-making process compared to deep machine learning (OECD, 2020).

Organizations refer to companies or entities that adopt or plan to deploy AI solutions in their operations (Personal Data Protection Commission Singapore, 2020).

Principles are rules of action that guide the development, procurement, and use of AI as well as the management of AI risks (Millar et al., 2018).

Risk is the magnitude of harm multiplied by the probability of such harm occurring (Olsen et al., 2009).