BOR version
23 August 2023

University of the Philippines
Principles for Responsible and Trustworthy Artificial Intelligence

Artificial Intelligence (AI) is the discipline concerned with the design and development of automated intelligent systems that perceive, reason out, formulate decisions, and act in an environment to achieve a set of measurable goals. AI systems embody computational structures that mimic human or animal cognition to process data, learn from experiences, and decide, plan, and act autonomously to satisfy a programmed objective.

In this document, AI is appreciated as machines that exhibit a certain level of human or animal intelligence, capable of problem-solving, decision-making, learning, and rational behavior. Further, AI is also a "socio-technical system" where "the processes used to develop [this] technology are more than their mathematical and computational constructs."[1] Fully understanding AI means taking into account "the values and behavior modeled from the datasets, the humans who interact with them, and the complex organizational factors that go into their commission, design, development, and ultimate deployment."[2]

Although the spread of AI provides excellent opportunities, it also creates significant risks.

AI makes lives easier by automating tasks and providing information and recommendations that suit individual needs. It is harnessed in making decisions on who gets a job, who is approved for a loan, what kind of medical treatment a patient receives, and what communities get policed.

AI can also be an essential tool for development. AI systems can revolutionize healthcare, transportation, and agriculture; aid in responding to climate issues; help in addressing poverty and hunger; and enhance personalized learning and improve education management. A study on AI and sustainable development goals (SDGs) published in 2020 revealed that "AI can enable the accomplishment of 134 targets across all the goals."[3]

However, the adoption of AI has led to increasing risks and hazards.

The *2023 AI Index Report* indicates that incidents of "ethical misuse of AI has increased 26 times since 2012."[4] Some experts are worried that people will misuse these systems to spread disinformation.

---

[1] Towards a Standard for Identifying and Managing Bias in Artificial Intelligence *NIST Special Publication 1270* March 2022 https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[2] Ibid
[3] Vinuesa, R., Azizpour, H., Leite, I. et al. "The role of artificial intelligence in achieving the Sustainable Development Goals" *Nature Communications* 11, 233 (2020) https://www.nature.com/articles/s41467-019-14108-y
[4] https://aiindex.stanford.edu/report/

Estimates show that AI deployment in the economy could also lead to massive job losses. The previously cited study on AI and SDG also reported that AI "may also inhibit 59 [SDG] targets."[5] In education, AI challenges include access for marginalized groups of students and privacy violations, such as unethical data collection and processing. Currently, many are worried that ChatGPT, Google Bard, and other generative AI applications open the door to cheating and plagiarism. There are also a few who fear that AI could slip out of human control.

The Philippines is committed to utilizing AI for development. The Department of Trade and Industry has developed an AI roadmap focusing on four areas: (1) digitization and infrastructure, (2) research and development, (3) workforce development, and (4) regulation. The Emerging Technology Development Division (ETDD) of the Department of Science and Technology-Philippine Council for Industry, Energy, Emerging Technology Research, and Development (DOST-PCIEERD) issued a report titled *Artificial Intelligence and Information & Communications Technology*. According to the secretary of the Department of Information and Communications Technology, "...government should come in and find ways to regulate it to ensure that AI is beneficial, that it is interoperable, it is transparent, and it is accountable."[6]

AI is seen to "make a significant contribution to the Philippine economy by 2030."[7] In terms of use, the *Generative AI Global Interest Report 2023* revealed that the Philippines has "the highest monthly search volume for AI tools overall: 5,052 per 100,000 population, mostly for text AI."[8] However, in *Government AI Readiness Index 2022*, referring to a government's readiness to use AI in delivering public services, the Philippines ranked 54th of 181 countries.[9] While it scored higher than the global average, it lags behind Singapore (2nd), Malaysia (29th), Thailand (31st), and Indonesia (43rd).

The University of the Philippines (UP) is actively engaged in developing AI in the country. AI is taught at the undergraduate and graduate levels. UP has the country's first Ph.D. program in AI, and UP faculty members and researchers are active in AI development. The UP Center for Intelligent Systems will conduct transdisciplinary research and education on artificial intelligence, data science, and complex systems.

For a national university that is committed to developing AI in the country, the challenge remains: how to promote positive and responsible use of AI and mitigate its negative consequences. It is therefore adopting the following *Principles for Responsible and Trustworthy AI* in order to provide guardrails and indicate the way forward on the development and use of AI in the University and the country.

It is also hoped that the adoption of these principles shall intensify the national discourse on the role of AI in national development.

---

[5] Vinuesa, The role of AI in achieving the SDG

[6] https://www.cnnphilippines.com/news/2023/6/15/dict-ai-regulation-workplace.html

[7] *The Economic Impact of Generative AI: The Future of Work in the Philippines* https://accesspartnership.com/the-economic-impact-of-generative-ai-the-future-of-work-in-the-philippines/

[8] https://www.electronicshub.org/generative-ai-global-interest-report-2023/

[9] Oxford Insight *Government AI Readiness Index 2022* available at https://www.unido.org/sites/default/files/files/2023-01/Government_AI_Readiness_2022_FV.pdf. In this index, Singapore is 2nd, Malaysia is 29th, Thailand is 31st, Indonesia is 43rd, and Vietnam is 55th.

Action of the Board of Regents
at its 1383RD Meeting on AUG 24 2023
APPROVAL
ROBERTO M. LARA
Secretary of the University
and of the Board of Regents

University of the Philippines
Principles for Responsible and Trustworthy Artificial Intelligence

1. **COMMON GOOD**. AI should benefit the Filipino people in particular, and humanity, in general by fostering inclusive economic growth, effective governance, sustainable development, and enhanced well-being while protecting the environment. AI systems should further the rule of law, human rights, and democracy.

2. **EMPOWERMENT**. AI should promote self-determination and bolster the capacity of humans to shape their future. Particularly, AI must empower vulnerable and marginalized groups.

3. **CULTURAL SENSITIVITY**. AI systems must be culturally responsive and culturally sustaining. Cultural norms, values, beliefs, and practices of users must be respected in designing, developing, and deploying AI systems.

4. **PRIVACY**. AI systems must incorporate privacy-by-design principles. Informed consent from users and maintaining the confidentiality of personal information must be upheld, when users provide information and when the system collects information about the users.

5. **ACCOUNTABILITY**. Individuals, groups, departments, institutes, colleges, and constituent universities involved in the development, deployment, and use of AI must take responsibility for the consequences of their actions. UP shall put into place mechanisms to hold the relevant stakeholders accountable for the AI system's actions and outcomes.

In Research and Development

6. **MEANINGFUL HUMAN CONTROL**. Humans should have decision-making authority over the AI's actions, decisions, and behaviors. AI systems should not operate in an unpredictable or unmanageable manner.

7. **TRANSPARENCY**. People should be able to understand how AI systems work. Individuals should be informed if AI-enabled tools are used. To the extent possible, the methods should be explainable. Finally, users should be able to understand AI-based outcomes and identify ways to seek remedies to harms that they may cause.

8. **FAIRNESS**. AI should be evaluated for gender bias, other forms of unfairness, and all forms of discrimination, especially in the data, models, and algorithms that are used. Collaborative procedures should be in place to actively identify, mitigate, and remedy these harms. AI developers should be mindful of its unintended consequences.

9. **SAFETY**. AI should never endanger lives. AI systems must function securely and safely. AI systems must be robust. In this context, robustness refers to the capacity of AI systems to

Action of the Board of Regents
at its 1383RD Meeting on AUG 24 2023
APPROVAL
ROBERTO M.J. LARA
Secretary of the University
and of the Board of Regents

endure and surmount adverse circumstances, including digital security threats. Compromising safety and security is unacceptable.

10. **ENVIRONMENT FRIENDLY**. AI should be evaluated in terms of its impacts on sustainability. AI models and tools must minimize risks to the environment. Developers should use computing resources more efficiently.

In Education

11. **PRIMACY OF LEARNING GOALS**. Decisions on the use of AI in teaching should start with the educational needs and priorities of learners. UP shall adopt AI systems that promote learner-centered pedagogy and foster collaborative and social learning. AI shall be used to improve the assessment of multiple dimensions of competencies and outcomes.

12. **HUMAN CAPITAL DEVELOPMENT**. UP shall strengthen existing programs and develop new ones to ensure that the country's AI workforce is highly skilled and ethical. These programs shall target women and other groups that are often excluded.

13. **CAPACITY BUILDING**. All members of the UP community must be AI literate. Additionally, faculty members must be trained in effectively using and integrating AI into teaching and learning practices. These two initiatives are necessary if faculty and students are to jointly innovate and benefit from the new technology as it further evolves.

14. **EDUCATION MANAGEMENT AND DELIVERY**. AI should improve university decision-making; make for more efficient administration, including admissions, enrollment, registration, human resource management, procurement, and inventory; and enable prompt regulatory compliance.

15. **COLLABORATION.** UP shall work with other universities, colleges, and research institutions to share best practices, co-develop AI courses and programs, undertake joint research and development, and advocate for responsible and trustworthy AI.

These *Principles for Responsible and Trustworthy Artificial Intelligence* shall serve as guardrails for our community and stakeholders.

Tensions are anticipated between these principles; hence, there is a need for policies, programs, and protocols that balance innovation and regulation.

Towards this, a multidisciplinary *UP AI Advancement Committee* (AIAC) is established.

The ultimate goal of the AIAC is to create an empowering environment where members of the UP community can continue to openly discuss the benefits and concerns associated with using AI and continue to come up with better policies and guidelines. This environment should also

Action of the Board of Regents
at its 1383RD Meeting on AUG 24 2023
APPROVAL
ROBERTO M. LARA
Secretary of the University
and of the Board of Regents

encourage the development and co-creation of AI tools among stakeholders. The AIAC shall also create a culture of cooperation, trust, and transparency among the university community, private sector, civil society, government, international organizations, and other stakeholders.

Immediately, the AIAC shall draft, for approval of the UP President, issuances to operationalize the development, deployment, and use of responsible and trustworthy AI. These issuances include, but are not limited to, the development of the following:

1. AI code of conduct based on the above-mentioned principles;
2. AI development program to accelerate AI research and development in the University;
3. Implementing framework for AI in UP education, including an AI literacy program, a faculty and staff upskilling program, and a research and teaching program on AI in society; and,
4. AI and digital infrastructure development to support the use of AI in teaching, research, and use. This includes providing equitable access to AI by members of the UP community

The UP President will appoint the chair and members of the AIAC and shall provide the necessary support for the committee to be able to discharge its functions.

A multi-sectoral *UP AI Advisory Board (AIAB)* is also created.

The AIAB shall advise the Board of Regents, through the UP President, on the following matters related to responsible and trustworthy AI:

1. Ensuring UP leadership in AI research, development, and teaching
2. Progressing UP's AI initiatives, managing and coordinating related activities, developing a resource leverage strategy, and enhancing opportunities for various sectors of society
3. Preparing the present and future Philippine workforce to use AI in all sectors of the economy
4. Coordinating ongoing AI research, development, and teaching activities among all agencies, higher educational institutions, and other organizations to ensure collaborations

The AIAB shall be composed of leaders from various sectors of society including, but not limited to, business / private sector, government / public sector, academia, and civil society.

The UP President shall appoint the chair and members of the AIAB. The AIAC shall provide secretariat support to the AIAB.

Action of the Board of Regents
at its 1383RD Meeting on AUG 24 2023
APPROVAL
ROBERTO M. LARA
Secretary of the University
and of the Board of Regents