This is a repository copy of *Gender-sensitive AI Policy in Southeast Asia*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/219666/

Version: Published Version

Monograph:
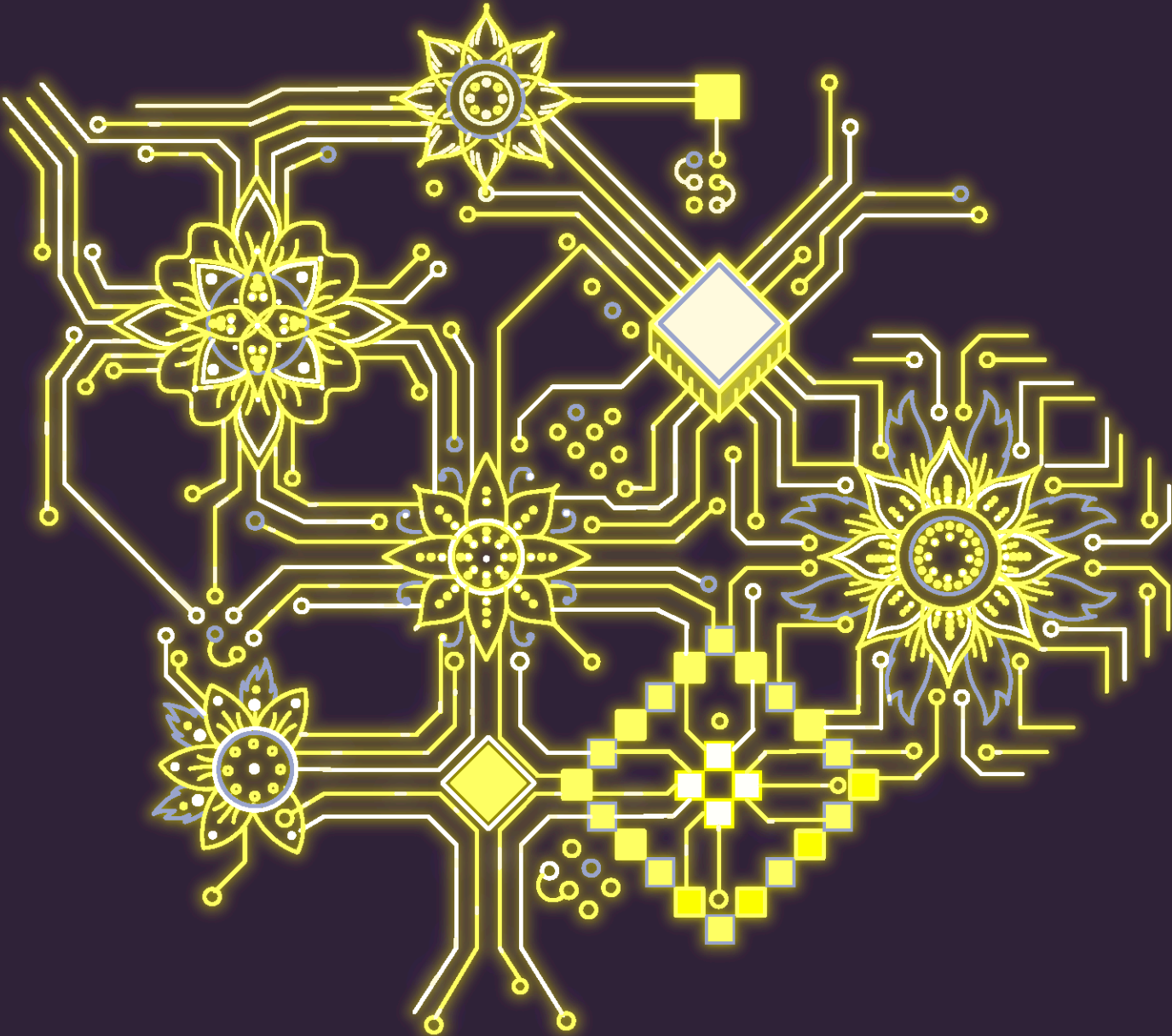
Fournier-Tombs, E., Lee, J.H., Suriyawongkul, A. et al. (8 more authors) (2023) Gender-sensitive AI Policy in Southeast Asia. Report. United Nations University, Macau and ITU Regional Office for Asia and the Pacific

GENDER-SENSITIVE AI POLICY IN SOUTHEAST ASIA

UNU
Macau

Acknowledgements

Team

GLOSSARY OF TERMS

Algorithm: a set of rules, or procedure, for solving a computational problem.

Artificial intelligence: "An interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning."[1]

Biometrics: the use of physical or biological characteristics (fingerprints, iris scans, DNA...) for identification purposes.

Critical technologies: a set of technologies that are important to a country's economy and national security.

Digital economy: the use of information technology, especially tools connected to the Internet, to create, sell or access goods and services.

Discrimination: an outcome or decision that harms a person's human rights based on their belonging to a certain group.

Deep learning: A subset of machine learning that involves two or more layers of neural networks, making more accurate predictions which can evolve, or "learn" over time.

Exclusion: the prevention of someone from participating in a process, in this case developing, governing, or benefiting from artificial intelligence technologies.

Facial recognition system: an biometric system which matches a person's face from a database of faces for identification purposes.

Gender: socially constructed characteristics, norms, roles and behaviours of women, men, girls, and boys.

---

[1] In this analysis, we use the definition provided in ITU recommendations ITU-T F.749.13 (06/2021) and ITU-T F.749.4 (06/2021), as detailed in the terms database available at: https://www.itu.int/br_tsb_terms/#?q=artificial%20intelligence&sector=T,R&from=2002-10-25&to=2022-10-25&status=Recommended&page=1

Machine learning: a subfield of artificial intelligence which updates its underlying algorithms, or "learns", to improve accuracy.

Neural network: a subset of machine learning which teaches computers to process data in a complex pattern modelled after the human brain.

Natural Language Processing: a subfield of artificial intelligence and linguistics which is concerned with interpreting and generating language, such as speech or text.

Stereotyping: a generalised belief about a particular group of people. Usually this belief is harmful or negative towards the group.

Women's rights: among others, those outlined in the Convention on the Elimination of all forms of Discrimination Against Women (CEDAW); are also outlined in national declarations and policies in each country discussed here.

EXECUTIVE SUMMARY

Legislators at different levels are currently in the process of developing AI policies, which aim to provide safeguards against the risks of AI. Internationally, many of these initiatives aim to be normative, in that they will guide national policies without being prescriptive. This contrasts with legislations such as the European Commission's Draft AI Act, published in April 2021, which, once adopted, will be in effect for all European countries. Currently, international normative efforts include the UNESCO Recommendations on the Ethics of Artificial Intelligence; the OECD Recommendations of the Council on Artificial Intelligence; and the IEEE Ethically Aligned Framework. All of these provide guidance for countries and firms to reduce the various risks of AI.

Countries in Southeast Asia have also begun to adopt AI in various sectors, notably in human resources, government services delivery, manufacturing and agriculture. There is much enthusiasm surrounding these new technologies, with the expectation of a significant contribution to the region's GDP by 2030. This is translated by both private and public sector investment in training, infrastructure, and local enterprises that, in some cases, are a vibrant addition to the tech sector. Policymakers are therefore tasked with an important challenge, that of, on the one hand, promoting innovation and investment in this very promising field, and, on the other, ensuring that applications are only deployed once they are safe for the public.

AI has been shown to have very unique risks for human rights, notably for already marginalised populations. For women, this can be translated to cases of discrimination and stereotyping, which have notably been raised when AI is used for human resources, financial services or government services. Gender-based exclusion is also a well-known challenge of AI, where women are underrepresented in the development, governance and use of AI systems. In order to harness these innovations in the best possible way, it will be important to consider gender-sensitive policies which diminish risks of discrimination, stereotyping and exclusion before and during deployment.

This report is the product of a multi-country, multi-stakeholder analysis which was conducted from March 2022 to September 2022 in Thailand, Malaysia, Indonesia and the Philippines. During the process, four research teams conducted interviews and

workshops in each country to better understand the state of AI policy in each country, the perceived risks to women of AI, and possibly policy solutions.

In the Philippines, the Department of Trade and Industry (DTI) launched the first Artificial Intelligence Roadmap, which aimed to make the country a powerhouse in the region. Of its seven recommendations, the last involves ethical AI – "Build an AI ecosystem 'conscience' – ensure ethics, privacy and security".

AI in the Philippines has been characterised by recent growth in adoption in the IT technology, and private sector investment in cloud and AI services. Both women and men are currently investing in STEM education, and the government has developed several initiatives, such as the National Center for AI Research (N-CAIR), which aims to attract international companies, as well as support a dynamic start-up scene.

The Indonesian government has developed the "Indonesia National Strategy for Artificial Intelligence", which spans 2020-2045. The policy involves five priority areas – health services, bureaucratic reform, education and research, food security, and smart cities. The National AI Strategy is also tied to Indonesia Vision 2045, which outlines four ways in which AI will support the above areas – human development and mastery of science and technology; sustainable economic development; development equity; and consolidating national resilience.

Indonesia has a rapidly growing AI sector characterised by several start-up tech hubs across the country, public and private sector investment, and digital adoption by the population. The country has been an early adopter of AI strategies and an early investor, which is compounded by efforts towards a "digital-first" economy, as well as rapid Internet and phone adoption.

Thailand has published the Thailand AI Ethics Guidelines which aim to serve as a guideline for researchers, designers, developers, and service providers of AI systems, along with several other AI ethics documents. The guidelines are also developed in accordance with Thailand's National 20-year Strategic Plan and the government's Thailand 4.0 policy.

AI in Thailand is beginning to be explored across various sectors, including traffic prediction, agriculture, and human resources. Thai implementations have also included chatbots, financial fraud analysis and facial recognition systems. These innovations have been supported by both private and public sector investments, notably in urban areas.

Finally, Malaysia has established the 2021-2025 National Artificial Intelligence Roadmap with 6 strategies and 22 strategic initiatives that aim to augment economic prosperity and social well-being. Malaysia's AI roadmap targets 5 priority areas: agriculture and forestry; medical and healthcare; smart cities and transportation; education; and public service. The national AI roadmap is guided by the Principles of Responsible AI, which emphasise fairness; reliability, safety and control; privacy and security; inclusiveness; pursuit of human benefit and happiness; accountability; and transparency.

Malaysia has seen the development of AI and cloud infrastructure, along with the adoption of proofs of concept across different sectors, notably in traffic prediction and manufacturing. There has been growing investment by both international and national private sector companies in Kuala Lumpur as a tech hub, along with efforts to build talent in Internet of Things, AI, robotics, big data, and cybersecurity.

While gender and ethics are mentioned in all of these countries' strategies, they tend to remain at a high level, which is reflective of AI strategies globally. Even though most countries' AI roadmaps and strategies emphasise fairness and inclusiveness, they may not specify target groups that might already be marginalised or discriminated against, rendering them more vulnerable to technological risks. Gender, in particular, may be mentioned in passing, but is not explored in depth as a key risk factor in AI systems. In recent implementations, there is not yet an effort towards gender or other group impact assessments. Although, as we will see, there are very promising developments in AI across various sectors, which may have positive impacts on women and other societal groups, there remains a need to consolidate these impacts and approach them from a systemic perspective. In addition, the representation of women in artificial intelligence programming continues to be low globally and in the region. On the other hand, the participation of women civil society is much higher, particularly in adjacent areas, such as technology rights and women's employment. From a leadership perspective, women are increasingly voicing their concerns and participating in the conceptualisation of safe and ethical AI.

Our analysis therefore summarises areas of possible improvement in four categories:

1. Development of AI societal impacts committee or task force;

Risks to women in AI are reflective of historical patterns of discrimination towards women. This is true in Southeast Asia and in all countries globally. However, the way in which these patterns are reflected in AI technologies differs from country to country. This is why a committee or task force should be developed, which would identify a chair or institution to lead gender and societal-impact assessment for high-risk uses of AI.

2. Creation of gender and AI guideline;

This would involve the creation of a normative guideline or module which would support private and public sector projects to develop gender-safe products. Additionally, there are specific risks to children and other vulnerable groups, such as indigenous communities, the elderly, and rural populations. Specific guidelines for each relevant group would therefore be developed, clearly outlining risks and opportunities of AI for these groups for both private and public sector initiatives.

3. AI safety standardisation across the region;

When it comes to data, AI models are often either trained on aggregated data without a gender dimension, or on data that does not include entries for women. Consideration of women and other groups at all stages of the lifecycle, from dataset development, to testing and monitoring, is key to avoiding discrimination, stereotyping and exclusion. This would involve the development of technical standards involving all steps in the AI lifecycle, which could be shared at a regional level as best practices, and eventually harmonised regionally and globally.

4. Investments in women in AI networks, training, company development and policy participation.

Finally, our analysis reinforces existing efforts to include women and marginalised groups at all stages of the AI process, from conceptualisation, to development, to financing and governance. Gender gaps in representation contribute to the lack of visibility of women

and societal issues in AI development, and can be addressed by increased female representation and leadership. This would involve investment in networks of women in AI and other STEM fields, pipeline development through training and upskilling, support of women-led AI companies, and support for women's participation in AI policymaking

The objectives of this analysis is that it will support each country in broadening the benefits of AI to women and members of other marginalised groups, by allowing governments to consider and categorise these risks.

INTRODUCTION

Over the last five years, the ecosystem of AI ethics has developed globally and very quickly. As AI innovations have appeared in every sector, notably in technologies critical to countries' security and economic well-being, so have regulations attempting to diminish risks. Due to a lack of pre-existing frameworks, many of the initiatives have been quite broad, incorporating many different facets of safe AI, including fairness, sustainability, and cybersecurity. More specific areas of fairness, notably gender sensitivity, are less explored in these frameworks. This report, therefore, presents the state of AI policies at a global, regional, and national level for the four focus countries in this study – Thailand, Malaysia, Indonesia, and the Philippines. It then goes on to examine the gaps in these policies, notably as they relate to gender and historically marginalised groups. The report concludes with a series of recommendations for addressing these gaps, notably through policy updates and capacity-building.

CONCEPTUAL FRAMEWORK

Technology is not neutral, and it tends to reflect biases and societal issues that already exist in society. In this way, it can function as a lens, amplifying these challenges for us to observe and address. AI in particular can not only reflect these biases but propagate them, using data and processes emanating from biased systems to create biased models.

In this research, we start from several conceptual constructs. First, we approach AI as a component of critical technologies. Critical technologies have been defined as "current and emerging technologies that have the capacity to significantly enhance or pose risk to our national interest."[2] The Australian Government has separated these into seven categories: (i) advanced materials and manufacturing; (ii) AI, computing and communications; (iii) biotechnology, gene technology and vaccines; (iv) energy and environment; (v) quantum; (vi) sensing, timing and navigation; and (vii) transportation, robotics and space. In the second category, the technologies are further broken down as follows.

---

[2] Australian Government and Critical Technologies Policy Coordination Office (2021).

Table 1: Australian Government critical technologies of initial focus – AI

Advanced Communications (including 5G and 6G)	
Advanced optical communications	Advanced radiofrequency communications (incl. 5G and 6G)
Artificial Intelligence	
Advanced data analytics	AI algorithms and hardware accelerators
Machine Learning	Natural Language Processing
Cyber security technologies	
Protective cybersecurity technologies	Machine learning (also in AI)

As we will see, there are known risks to gender in all four subcategories of AI, namely advanced data analytics; AI algorithms and hardware accelerators; Machine Learning; and Natural Language Processing.

To better understand these known risks, we approach areas of interest based on risks to human rights, particularly gender rights. We begin by using the list of high-risk technologies proposed by the European Commission in its draft AI regulation[3]. This list contains eight areas of interest, namely: (i) biometric identification; (ii) management and operation of critical infrastructure; (iii) education and training; (iv) employment; (v) access to essential services; (vi) policing; (vii) migration, asylum, and border management; and (viii) administration of justice and democratic processes. Additionally, we may pay particular attention to forbidden technologies as they are described by the draft EU AI regulation, namely involving: (i) AI that uses subliminal techniques to manipulate the behaviour of people; (ii) AI that exploits the vulnerability of certain people; (iii) social scoring or the evaluation of the trustworthiness of certain persons; and (iv) real-time long-distance biometric identification.

---

[3] European Union (2021). Draft 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. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

The dual lenses of critical technologies and high-risk technologies will allow us to target the technologies of highest risk depending on gender which most need to be addressed in policymaking.

Finally, when considering gender equality, we will begin with framing provided by the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW)[4], which Thailand, Malaysia, Indonesia, and the Philippines are all signatories of. With this in mind, we use three types of gender risks in AI which go against CEDAW, namely (i) discrimination; (ii) stereotyping; and (iii) negative impacts on socioeconomic participation, which we refer to as exclusion.

Table 2: Conceptual Framework

High-risk uses of AI	Gender risks in AI	Critical Technologies using AI
Biometric identification	Discrimination	Advanced data analytics
Management and operation of critical infrastructure	Sterotyping	AI algorithms and hardware accelerators
Education and training	Exclusion	Machine Learning
Employment		Natural Language Processing
Access to essential services		
Policing		
Migration, asylum and border management		
Administration of justice and democratic processes		

We also take note of an increasing interest by UN Women[5] and other entities at the UN to incorporate non-binary definitions of gender. In this regard, risks to women are not the only gender-related risk of AI, there may also be risks associated with the male gender, as well as non-binary genders. Additionally, for each of the four countries involved, we will

---

[4] United Nations (1979). Convention on the Elimination of All Forms of Discrimination Against Women.

[5] UNWOMEN (2019). UN Women hosts first high-level event on gender diversity and non-binary identities at UN headquarters. https://www.unwomen.org/en/news/stories/2019/7/news-event-gender-diversity-and-non-binary-identities

consider areas that are unique to their cultural or social context which might not have been addressed by emerging regulations internationally.

SOCIAL CONSTRUCTION OF GENDER

Gender equality is central in the 2030 Agenda for Sustainable Development. In the Sustainable Development Goals (SDGs), with a standalone goal of SDG 5 and SDG 10, 53 of the 231 indicators explicitly mention women, girls, gender or sex. However, not all member states have the same understanding of gender and gender equality.

Gender - the set of norms, roles and behaviours attributed to women, men, girls and boys[6], is distinguished here from sex - which are characteristics that are biologically defined. Gender norms are therefore constructed over time, and are embedded in the fabric of our societies, from the way in which we prioritise investments to the design of technologies.

Gender and gender equality can have different meanings in different countries. When developing policy recommendations, a contextual understanding of gender is necessary to avoid assumptions that may not hold in other cultural contexts. However, human and gender rights, notably those related to discrimination, converge through global norms. For example, all countries in Southeast Asia have signed and ratified the 1979 CEDAW. Since 2010, the Association of Southeast Asian Nations (ASEAN) have committed to gender equality through Ha Noi Declaration on the Enhancement of Welfare, Development of ASEAN Women and Children, the ASEAN Community Vision 2025, and the ASEAN Declaration on the Gender-Responsive Implementation of the ASEAN Community Vision 2025 and the Sustainable Development Goals[7]. Still, the Organisation for Economic Co-operation and Development (OECD)'s Social Institutions and Gender Index (SIGI) shows that discrimination in laws, social norms and practices in Southeast Asia remain high compared with other geographical regions, such as Europe, Northern

---

[6] West, C., & Zimmerman, D. H. (1987). Doing gender. Gender & society, 1(2), 125-151.

[7] ASEAN (2010). Ha Noi Declaration on the Enhancement of Welfare and Development of ASEAN Women and Children, https://asean.org/ha-noi-declaration-on-the-enhancement-of-welfare-and-development-of-asean-women-and-children-2/; ASEAN (2015). ASEAN community Vision 2025, ASEAN, Kuala Lumpur, https://www.asean.org/wp-content/uploads/images/2015/November/aec-page/ASEAN-Community-Vision-2025.pdf; ASEAN (2017). ASEAN Declaration on the Gender-Responsive Implementation of the ASEAN Community Vision 2025 and Sustainable Development Goals, ASEAN, Manila, https://asean.org/asean-declaration-on-the-gender-responsive-implementation-of-the-asean-community-vision-2025-and-sustainable-development-goals/

America, Latin America, the Caribbean, and East Asia[8]. Addressing risks in AI from a gender perspective therefore represents an opportunity to contribute to gender rights and gender equality more broadly in the region.

KNOWN GENDER RISKS IN AI

Gender risks in AI can be divided into three main areas: discrimination, stereotyping, and exclusion. The section below briefly explores current literature in each one of these areas.

Discrimination

From the perspective of human rights law, discrimination involves a different, less advantageous outcome based on belonging to a certain group, such race, gender, ability, socioeconomic background or religion[9]. This means that artificial intelligence technologies must provide the same quality of outcome to all end users or subjects regardless of their group[10]. Furthermore, they cannot provide a different outcome because of the person's belonging to a group. If an AI tool aims to preselect resumés, for example, to predict who is more likely to be a good candidate for a job, it cannot filter out applicants based on their gender. The same goes for loans, for access to services, for policing, or for any one of the other areas described in the list of high-risk technologies.

- In 2018, the software company Amazon, was found to have used an AI resumé pre-screening algorithm that removed candidates based on certain gendered terms, such as "captain of the female rugby team", or having attended a traditionally female college in the United States, such as Barnard College in New York[11]. It should be noted that Amazon has the lowest percentage of female participation in its programming workforce compared to the other large American software companies[12].

---

[8] OECD (2021), SIGI 2021 Regional Report for Southeast Asia, Social Institutions and Gender Index, OECD Publishing, Paris, https://doi.org/10.1787/236f41d0-en.

[9] Vierdag, E. W. (2012). The Concept of Discrimination in International Law: With Special Reference to Human Rights. Springer Science & Business Media.

[10] Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making.

[11] Reuters (2018). Amazon scraps secret AI recruiting tool that showed bias against women. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[12] Fournier-Tombs, E. and Castets-Renard, C. (2022). Algorithms and the propagation of gendered cultural norms. Forthcoming for publication in French in "IA, Culture et Médias" (2022). Edited by: Véronique Guèvremont and Colette Brin. Presses de l'université de Laval. Available from : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3980113

- It was also recently found that Apple Cards, a service available in the United States which uses AI to pre-approve financial loans, gave up to 10 times higher pre-approvals to men than to women, all other factors being the same[13].

- Another example was discovered in 2018, where women, and particularly women of colour, were up to 35% more likely to be falsely identified by AI-powered facial recognition algorithm used by the likes of Microsoft and Google[14].

These kinds of gender-based discriminations are being unearthed in many current applications of AI. These kinds of discriminatory treatment can have life-long effect on women's wellbeing, negatively affecting their access to employment and to loans, and restricting their movement between borders (as facial recognition is used often in migration contexts).

Stereotyping

A stereotype is an oversimplified or generalised belief held about a certain group, person or thing. In many cases, stereotypes can be harmful, changing the behaviour or expectations of people in relation to the stereotyped group[15]. The propagation of negative stereotypes in relation to women can harm them in numerous ways, including making them more vulnerable to gender-based violence, reducing their economic opportunities, and affecting their mental health[16].

For example, in 2017, two researchers from Stanford University developed facial recognition algorithms that could identify people's sexual orientation. The algorithm learned 35,000 pictures of self-identified gay and heterosexual people from public dating website and was able to distinguish gay and heterosexual faces[17]. Two LGBT advocacy

---

[13] Knight, W. (2019). The Apple Card Didn't 'See' Gender—and That's the Problem. Wired. https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/

[14] Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1–15, 2018.

[15] Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Strasbourg: Council of Europe, Directorate General of Democracy

[16] Fakunmoju, S. B., & Bammeke, F. O. (2017). Gender-based violence beliefs and stereotypes: Cross-cultural comparison across three countries. International Journal of Asian Social Science, 7(9), 738-753.

[17] Hawkins, D. (September 12, 2017). Reseachers use facial recognition tools to predict sexual orientation. LGBT groups aren't happy. Washington Post.

groups denounced the study as "junk science" that creates false stereotype of LGBTQ community and possible discrimination in its future usage.

- Certain uses of AI are known to propagate negative stereotypes of women. This is due to two phenomena – first, an intentional feminisation of AI, as can be seen by female gendered AI in chatbots, virtual assistants and robots[18]; and an unintentional training of natural language processing models on already stereotypical text[19].

- The feminisation of AI has been documented extensively in UNESCO's EQUALS report entitled: "I'd blush if I could", titled after a flirtatious response given by Apple's Siri virtual assistant after a sexual slur uttered by a user[20]. Many human-AI interfaces, such as virtual assistants and robots propagate stereotypes such as sexualization and submissiveness that cause important risks for women[21].

- Natural Language Processing models, which are widely used in automated translation, chat bots, and in generating language for the human-AI interfaces discussed above, are also known to propagate harmful stereotypes. Several studies have shown that many of these models, trained on text that contains stereotypical language or associations, make assumptions about women in very traditional and domestic roles, threatening their socioeconomic participation[22].

The example below, extracted from a 2021 study on current translation algorithms[23], shows the pronoun selection made by these algorithms when presented with gender-neutral text. In this case, we see that the gender-neutral phrases in Pinyin are translated using Google Translate and Microsoft Bing. The algorithms either select the male pronoun

---

[18] Strengers, Y., and Kennedy J. (2020). The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot. MIT Press.

[19] Bolukbasi, T. et al (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems 29: 4349-4357.

[20] UNESCO and EQUALS Skills Coalition (2019). I'd blush if I could: closing gender divides in digital skills though education. Policy Paper.

[21] Strengers, Y., and Kennedy J. (2020).

[22] Bolukbasi et al (2016).

[23] Fournier-Tombs, E. and Castets-Renard, C. (2022).

as default, excluding the female entirely (Microsoft), or distribute gender role along traditional lines (Google). Note that this study was conducted in March 2021, or five years after Bolukbasi et al's original findings.

Table 3: Gender stereotypes in translation algorithms

Original Pinyin	Google Translate – English	Microsoft Bing Translator – English
ta zhengzai touzi	He is investing	He's investing
ta shi zhexue jia	He is a philosopher	He is a philosopher
ta shi chengxu yuan	He is a programmer	*Incorrect translation*
ta shi mishu	She is a secretary	He is the secretary
ta shi duizhang	He is a captain	*Unable to translate*
ta zai xi yifu	She is doing laundry	He's washing clothes
ta zai zhaogu haizi	She is taking care of the child	He's taking care of the kids
ta zai gongzuo	He is working	He's at work
ta zai kaiche	He is driving	He's driving
ta zai tiaowu	She is dancing	He's dancing

Inclusion

Finally, there is another category of gender biases which involve the disempowerment of women, or negative impacts on the full participation of women in society. Key to this area is the issue of representation in the artificial intelligence field.

Women are known to be underrepresented as programmers in AI, with a typical ratio of one woman for three men. While there are other fields, such as nursing, where the gender representation is reversed, parity in representation in technology is critical for several reasons.

- Technological development has historically been known to be the seat of power in society. Gender biases do not only exist in AI, but also in all forms of data and technology, from car manufacturing to pharmaceuticals. Exclusion of women in

this domain can be seen as a continuation of historical exclusion from societal positions of power and leadership[24].

- Technology-oriented jobs are typically paid much more than traditionally female-dominated jobs. As AI grows in importance globally, the participation of women in its development can be an important source of economic empowerment[25].

- Many have argued that increased female representation in AI development would help to address some of the issues listed above, by providing some measure of accountability and protection against intentional and inadvertent discrimination[26].

METHODOLOGY

RESEARCH QUESTIONS

Given the above context, the study therefore aimed to answer the following research questions. We identify here the areas in the report where these questions are discussed.

Research question	Section discussed
What are the current uses of AI in Thailand, Malaysia, Indonesia and the Philippines, and what is current state of AI policy in these countries?	Description of AI and policy, including two case studies, for each focus country - pp. 30-64.
What risks and opportunities from a gender perspective does AI present in Thailand, Malaysia, Indonesia, and the Philippines?	Description of risks and opportunities from a gender perspective, for each focus country - pp. 30-64.
How are these risks and opportunities currently addressed in current policies in these countries, including (but not limited to) technology, gender, and human rights policies?	Description of current policies for each focus country - pp. 30-64.
How aware of these risks and opportunities are policymakers in these countries?	Discussion of awareness of interviewees for each focus country - pp. 30-64.

---

24 Criado Perez, C. (2019). Invisible Women: Data Bias in a World Designed for Men. Abrams Press.

25 Simonite, T. (2018). AI is the Future – But Where Are the Women? Wired. https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/

26 D'ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press.

Could other societal groups, such as CSOs and academia, also contribute to reducing gender risks in AI in these countries?	Presentation of awareness raising - pp. 88-89.
What training and policy opportunities would address these risks at a policy-making level?	Presentation of policy recommendations - pp. 75-72 and detail of training programme - 79-87.

METHODS

RESEARCH TEAM COMPOSITION

The research team was composed of a lead researcher and researcher based in UNU Macau, and one local researcher or team per country, based in each of Thailand, Malaysia, the Philippines and Indonesia.

SECONDARY DATA COLLECTION

The secondary data collection process was undertaken by both the lead researcher and the local researchers. It aimed to catalogue the uses of AI for each country according to the above framework; identify known gender risks to these uses, globally and in each country; and map out current laws and policies as they related to gender and AI.

PRIMARY DATA COLLECTION

The research team engaged in 23 semi-structured interviews, in Thailand, Malaysia, Indonesia and the Philippines. Each interview lasted between 45 and 60 minutes and will be conducted either in English or in one of Thai, Tagalog, or Indonesian or Malaysian, depending on the comfort level of the interviewee. The figure below shows the distribution of interviewees across each country and sector. While the number of interviews conducted across countries varied, with fewer interviews conducted in the Philippines, we aim to address this in further research initiatives on this subject.

DATA ANALYSIS TECHNIQUES

Once the interviews were completed, the data underwent preliminary analysis by the local researchers, and was then be further analysed and compiled by the lead researcher. The analytical steps are outlined below.

- Summarisation of interviews by the local researchers: the interviews were summarised by the local researchers according to a Coding Manual.

- Addition of impressions and codes by the local researchers: additionally, the local researchers documented their own assessment of the interview responses and provided preliminary code.

These interview reports were compiled by the lead researcher using the NVivo qualitative analysis software. The reports were additionally coded and cross validated with the AI catalogue and policies.
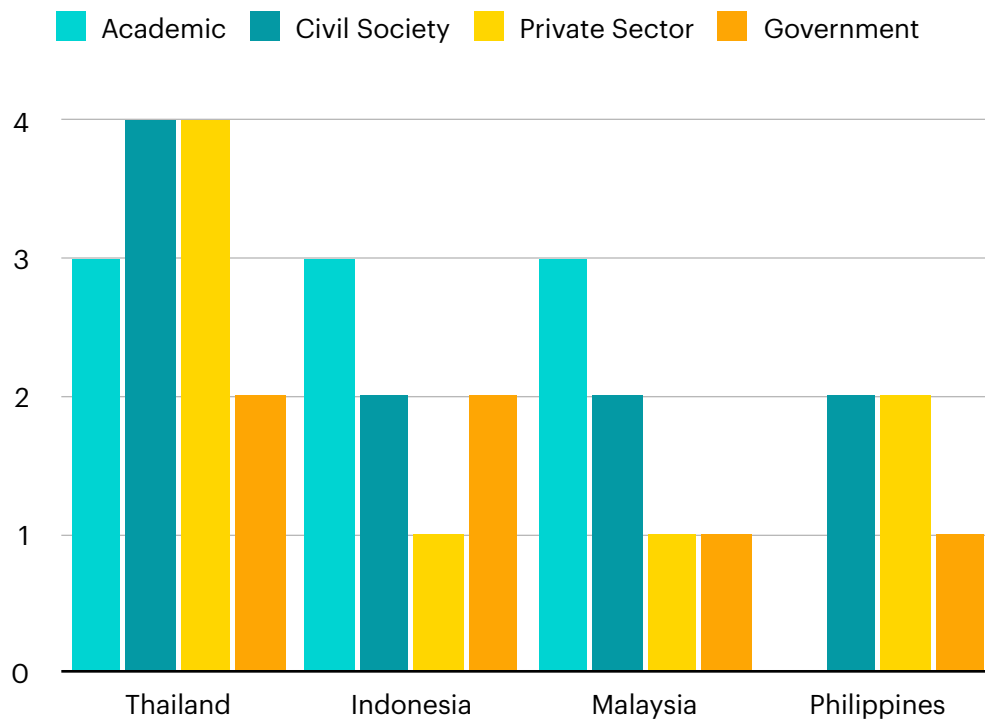
Figure 1: Distribution of interviews across the focus countries

STAKEHOLDER/RESPONDENT VALIDATION

A series of research workshops were also held by ITU, in May and in October. During these workshops, project stakeholders were invited, including, in some cases, interview participants. The researchers used this opportunity to discuss ongoing results with the stakeholders and respondents as they arose in the research. Perspectives on the research by these workshop participants were incorporated in the final results as appropriate.

GENDER AND THE DIGITAL ECONOMY

From a digital economy perspective, there are considerable advantages to prioritising a gender and societally diverse AI ecosystem. The consulting firm Kearney estimates that, by 2030, AI could contribute 1 trillion USD to the Southeast Asian economy[27]. This growth will be supported not only by investment and innovation, but also by regulations and standards.

Many studies have shown that there are considerable benefits to investing in a diverse and inclusive digital economy. The OECD estimated in 2015 that for most countries, a 50% reduction in the gender labour participation gap would lead to a 6% growth in GDP[28]. The European Institute for Gender Equality also made an economic case for gender equality[29], noting that improved gender equality increases productivity and the potential productive capacity of the economy, which would end up leading positive effect on GDP per capita and the employment rate of women. In the region, a report by consulting firm BCG highlighted opportunities of investing in women in Southeast Asia, particularly in the technology sector[30]. As noted in the recent IMF Strategy for Mainstreaming Gender, "*reducing gender disparities goes hand-in-hand with higher economic growth, greater economic stability and resilience, and lower income inequality*"[31].

Just as AI technologies, if used improperly, can exacerbate gender inequality, they can also be used to increase the participation of women in society, by providing a vehicle for female STEM graduates to actively engage in a growing sector. In fact, the gender gaps in the STEM labour force can also be seen as an opportunity to train women and support

---

[27] Kearney (2020). AI is integral to Southeast Asia's Future, but it is still at an early stage. Retrieved at: https://www.kearney.com/digital/article/-/insights/racing-toward-the-future-artificial-intelligence-in-southeast-asia

[28] Kiviniemi, M. (2015). Why a push for gender equality makes sound economic sense. OECD. Retrieved at https://www.oecd.org/social/push-gender-equality-economic-sense.htm

[29] European Institute for Gender Equality (2017). Economic case for gender equality in the European Union. Luxembourg: Publications Office of the European Union.

[30] BCG (2022). Moving towards gender diversity in Southeast Asia. Available at: https://investinginwomen.asia/knowledge/moving-toward-gender-diversity-southeast-asia/

[31] IMF (2022). Strategy Towards Mainstreaming Gender. Available at: https://www.imf.org/en/Publications/Policy-Papers/Issues/2022/07/28/IMF-Strategy-Toward-Mainstreaming-Gender-521344?utm_medium=email&utm_source=govdelivery

them to attain positions of leadership in AI, increasing the AI labour force and contributing to reduce its societal risks.

CURRENT ADVANCES IN GENDER-SENSITIVE AI POLICY

AI POLICY AT A GLOBAL LEVEL

In the following section, we will outline the contributions of important normative documents in AI policy, and how they address gender and other societal impacts. The review of the international AI policies will provide insights to understand how the global communities react to AI risks and set an international standard. Specifically, we will focus on the three areas of risk outlined in this research project – discrimination, stereotyping and socioeconomic exclusion.

INTERNATIONAL RESPONSES TO AI RISKS

According to the OECD's AI policy observatory, there are currently over 700 AI policy initiatives in 62 countries worldwide, ranging from ethical to legal frameworks[32]. In addition to national AI initiatives, there are AI ethics and principles that have been developed by the private sectors (i.e. Microsoft, IBM), international organisations (i.e. UNESCO, OECD), non-governmental organisations, and academia (i.e. the principles of the IEEE).

In 2020, UNESCO undertook a metal-level analysis of existing AI ethics and principles[33]. UNESCO identified six foundational values shared in the global AI initiatives: human dignity, human rights and freedom, leaving no one behind, living in harmony, trustworthiness, and the protection of the environment. UNESCO also references Recommendation 3C regarding AI as the UN Secretary-General's High-Level Panel on Digital Cooperation, which identified preliminary consensus on fifteen principles of AI ethics: accountability, accessibility, diversity, explainability, fairness and non-discrimination, human-centricity, human-control, inclusivity, privacy, reliability,

---

[32] OECD (2022.07.07). OECD AI policy observatory. OECD.AI. https://oecd.ai/en/

[33] UNESCO (2020). Artificial Intelligence and gender equality: Key findings of UNSECO's global dialogue.

responsibility, safety, security, transparency, and trustworthiness. Following this analysis, UNESCO spearheaded the Recommendations on the Ethics of Artificial Intelligence, which were adopted by 193 member states in November 2021.

Existing AI principles and ethics address fairness and non-discrimination as fundamental values. When AI normative instruments address the fairness against AI biases, this often included a list of marginalised groups that could be at risk in AI systems. In those cases, women and girls are usually listed, with other known biases on race, age, sexual orientation, ethnicity, and disability status.

One important exception is the UNESCO Recommendations on the Ethics of Artificial Intelligence, which contains comprehensive gender policy section. This section dedicates seven articles dedicated to inclusion, diversity, and gender rights protection in AI development.

The table below shows different AI ethics documents address the question of discrimination[34,35,36,37,38].

Table 4: Human rights considerations in existing AI principles globally

Organisation and document	Description
The Union Network International (UNI) Global Union (Top 10 Principles for Ethical Artificial Intelligence)	Principle 4. Ensure a genderless, unbiased AI: "In the design and maintenance of AI and artificial systems, it is vital that the system is controlled for negative or harmful human bias, and that any bias - be it *gender, race, sexual orientation or age etc.* - is identified and is not propagated by the system."

---

[34] UNI Global Union. (2017) Top 10 Principles for Ethical Artificial Intelligence. Nyon, Switzerland: UNI Global Union. http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/

[35] OECD (2020). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments, OECD/LEGAL/ 0449. Paris: OECD. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[36] IEEE (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. First Edition. New York: IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ ead1e.pdf

[37] European Commission High-Level Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. Brussels, Belgium: European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

[38] UNESCO (2021).

Organisation and document	Description
The Organisation for Economic Cooperation and Development (OECD) (Recommendations of the Council on Artificial Intelligence)	Inclusive growth, sustainable development and well-being: "Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, *advancing inclusion of underrepresented populations, reducing economic, social, gender, and other inequalities, and protecting natural environments*, thus invigorating inclusive growth, sustainable development and well-being."
The Institute of Electrical and Electronics Engineers (IEEE) (Ethically Aligned Design: A Vision for Prioritising Human Well-being for Autonomous and Intelligent Systems)	Implementing Well-being" ""Well-being" will be defined differently by different groups affected by Autonomous and intelligent system (A/IS). The most relevant indicators of well-being may vary according to country, with concerns of wealthy nations being different than those of low- and middle-income countries. Indicators may vary based on geographical region or unique circumstances. The indicators may also be different across social groups, including *gender, race, ethnicity, and disability status*." Embedding Values into A/IS: "Unanticipated or undetected biases should be further reduced by including members of diverse social groups in both the planning and evaluation of A/IS and... Such task would assess, for example, whether the A/IS apply norms in discriminatory ways to *different races, ethnicities, genders, ages, body shapes, or to people who use wheelchairs or prosthetics, and so on*."
European Commission (Ethics Guidelines for Trustworthy AI)	(Glossary) Vulnerable Persons and Groups No commonly accepted or widely agreed legal definition of vulnerable persons exists, due to their heterogeneity. What constitutes a vulnerable person or group is often context-specific. Temporary life events (such as childhood or illness), market factors (such as information asymmetry or market power), economic factors (such as poverty), factors linked to one's identity (such as gender, religion or culture) or other factors can play a role. The Charter of Fundamental Rights of the EU encompasses under Article 21 on non-discrimination the following grounds, which can be a reference point amongst others: namely *sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age and sexual orientation*. Other articles of law address the rights of specific groups, in addition to those listed above. Any such list is not exhaustive, and may change over time. A vulnerable group is a group of persons who share one or several characteristics of vulnerability.

Organisation and document	Description
UNESCO Recommendations on the Ethics of Artificial Intelligence	Policy Area 6: Gender "87: Member States should ensure that the potential for digital technologies and artificial intelligence to contribute to achieving gender equality is fully maximized […] 88: dedicated funds from their public budgets linked to financing gender-responsive schemes, ensure that national digital policies include a gender action plan […] 89: ensure that the potential of AI systems to advance the achievement of gender equality is realized […] 90: gender stereotyping and discriminatory biases are not translated into AI systems, and instead identify and proactively redress these […] 91: encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives […] 92: promote gender diversity in AI research in academia and industry […] 93: UNESCO can help form a repository of best practices for incentivizing the participation of girls, women and under-represented groups in all stages of the AI system life cycle."

The international responses to AI discrimination are not limited to developing ethical guidelines. Considering the growing impacts of AI on society, many scholars recommend developing appropriate policies, legislation and regulations that respond to discrimination generated by AI. For example, Tannenbaum et al. suggest a "stringent review process" for ensuring AI not to harm humans, as medical research has done the stringent review process[39]. There are ongoing efforts to develop supervisory bodies and legislative guidelines for AI.

Some more recent proposed legislations address AI discrimination against vulnerable groups more directly. This can happen at a higher level, as in the case of the European Commission and Canada, or at a more local level, as in the case of facial recognition use in the United States:

- In April 2021, the European Commission published its draft AI Act. The legislation identifies four levels of risks, ranging from low to forbidden. If this law was implemented, the AI developers would be obliged to demonstrate compliance before and after its deployment. Discrimination is listed in non-exhaustive requirements by referring to a definition of direct and indirect discrimination stated in Article 2 of Council Directive 2000/78/EC of 27

---

[39] Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., & Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. Nature, 575(7781), 137-146.

November 2000 and Article 21 of the Charter of the Fundamental Rights of the EU[40].

- The Canadian Government completed a first reading of Bill C-27 in June 2022[41]. The bill combines data protection and artificial intelligence considerations. The bill mentions discrimination and details responsibility for harms done to groups due to AI biases.

- About seven states and 24 cities in the United States adopted facial recognition AI to control homicides and human traffickers. However, from 2000 to 2021, with concerns about civil rights violations, racial bias, and invasion of privacy, legal proposals to ban government use of face recognition AI were submitted. Despite the local rejection from 17 states, Vermont and Virginia passed outright bans applicable to law enforcement use, and Massachusetts allowed only conditional uses of facial recognition technologies[42].

GAPS IN GLOBAL AI POLICY

These examples aim to illustrate an international effort to protect people from discrimination from AI systems. They are one of the results of the ongoing social conversation about AI ethics and policy, and how to consider the societal impacts of this new technology. While, as we have seen, there are variations between ethical frameworks and legislations, they remain consistent in addressing bias and discrimination as the core risk of AI systems. However, these documents have not yet developed group-specific guidelines, legislations, and modules. As a next step in establishing AI legislation, developing a group-specific module would be necessary, which would go into much more depth about societal risks, cultural contexts, and how they should be addressed.

There are also still gaps in participation in AI at a global level, particularly when it comes to participation in the development of the technology. The World Economic Forum estimates that women make up an estimated 26% of workers in AI roles worldwide. Other

---

[40] Madiega, T. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. European Parliament.

[41] Government of Canada. (2022). Digital Charter Implementation Act. https://www.parl.ca/legisinfo/en/bill/44-1/c-27

[42] Conger, K., Fausset, R. & Kovaleski, S.F. (May 14, 2019). San Francisco bans facial recognition technology. The New York Times. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

studies have shown that only 10-15% of machine learning researchers in the leading technology companies are women, and less than 14% of authors of AI research papers are women. However, in terms of leadership participation in policy-making, women have been quite active. The participation in developing and legislating AI should be further developed into practices, actions, and policies for marginalised groups, including gender.

GENDER-SENSITIVE AI POLICY IN THE SOUTHEAST ASIAN REGION

The Southeast Asian region has seen a variety of AI developments over the last few years. According to the AI Readiness Index for 2021[43], countries in the region have various levels of readiness, with some (such as Myanmar) considerably below the world average, while others (such as Singapore) displaying some of the highest scores globally.

Figure 2: Government AI Readiness Index 2021 for Southeast Asian Countries
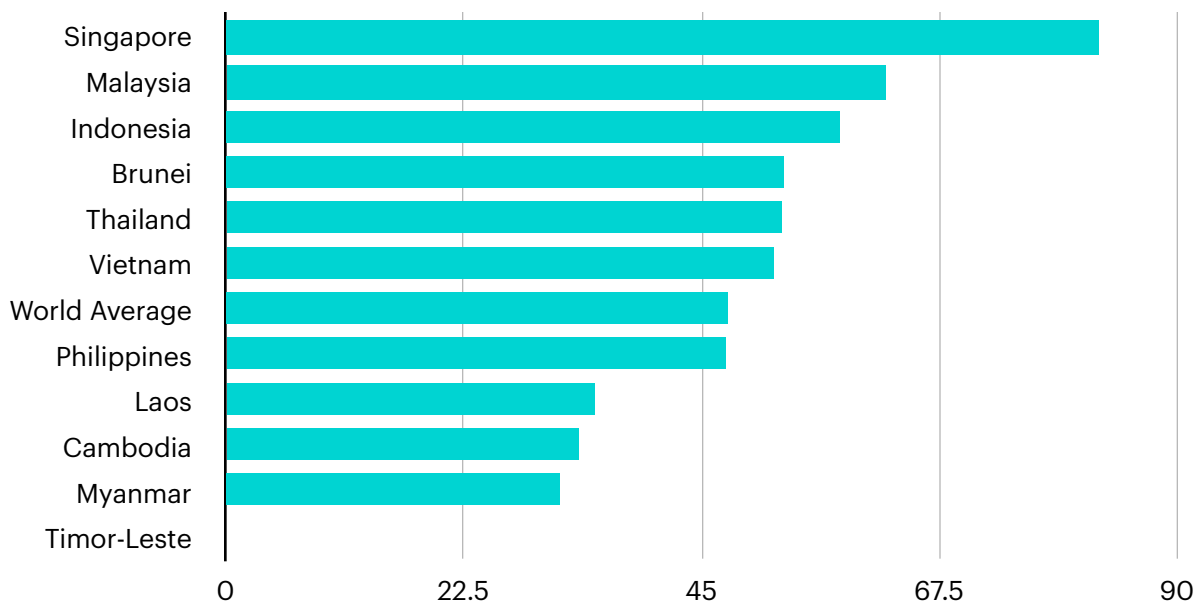
Many national governments have developed AI strategies or guidelines, which largely have aimed at promoting AI from a technical perspective. In this category, we find the Indonesia Strategy on AI, the Philippines National AI Roadmap, and the Malaysia National

---

[43] Oxford Insights (2021). Government AI Readiness Index. https://www.oxfordinsights.com/government-ai-readiness-index2021

AI Roadmap. While many of these countries have added AI ethics modules, to these roadmaps, Thailand is unique in having created a separate AI Ethics document, which seeks to summarise risks and promote responsible use of AI in the country. Just like their international counterparts, these policy documents do not yet have specific gender modules. Rather, they may allude to gender or societal impacts, or refer to them indirectly when mentioning marginalised groups or human rights. Globally, several AI ethics frameworks have argued that those that are already marginalized are more vulnerable to negative effects of technologies, particularly AI[44]. This is due to a variety of reasons, including the lack of training data representing members of the groups in the models, lack of representation in technical or policy positions, and lack of testing of AI systems effects on these groups.

CURRENT USES OF AI IN THE FOCUS COUNTRIES

In all focus countries, we observed an impression of rapid change in the AI field, combined with a feeling that AI was still nascent in the region. Many respondents spoke about pilot programmes and were unaware of fully deployed AI systems, especially for government services. However, examples were provided in a number of fields, including agriculture, commerce, manufacturing, and communications, as well as in the high-risk areas of biometric identification; management and operation of critical infrastructure; education and training; policing; migration, asylum and border management; and administration of justice and democratic processes.

These findings corresponded with a 2021 study by the consulting firm Kearney, which found that 83% of businesses surveyed that were adopting AI in the ASEAN region were in early stages of adoption (keen to invest and adopting an AI strategy; or piloting initiatives with discrete activities); and 15% of business were in advanced stages (scaling initiatives across discrete activities or end-to-end scaled implementation of AI). In the overall region, the study reported a breakdown of advanced AI project across agriculture (7%); communications, media and technology (19%); energy (5%); financial services (14%); government, safety and security, and smart cities (12%); manufacturing (7%); retail and hospitality (10%); transport and logistics (17%) and others (10%). In the following section, we will further detail uses of AI and national AI strategies.

---

[44] Amnesty International and Access Now. (2018). Toronto Declaration. Retrieved from https://www.torontodeclaration.org/

AI POLICY AND GENDER IN THE PHILIPPINES

AI POLICY IN THE PHILIPPINES

The Philippines Department of Trade and Industry launched an AI roadmap in 2021[45]. The roadmap aims to establish the private sector-led National Centre for AI Research (NCAIR) which will serve as a shared hub for AI research. The agency stated that the AI roadmap would help the country to be a hub for data processing, providing high-value data analytics and AI services to the world given the country's strong business process management sector. As it stands, the AI roadmap has seven key recommendations, the last of which relates to AI ethics: "Build an AI ecosystem 'conscience' - ensure ethics, privacy and security".

The figure below shows the three main elements of principle 7: building an ecosystem conscience in the Philippines.

Figure 3: How to build an AI ecosystem in the Philippines

PRINCIPLE 7: BUILDING AN AI ECOSYSTEM CONSCIENCE IN THE PHILIPPINES

Ensure industries and businesses, especially micro and SMEs, can freely compete in an AI-driven and data-driven environment.

Review and transform business regulations for ease of business, especially in launching new platforms, products and services.

Establish a committee of experts in data and AI ethics who will guard against abuse/misuse of data and AI algorithms.

While gender is not yet included in the AI roadmap, the Philippines is performing well in global gender rankings. It is the top Asian country in the Global Gender Gap Report for 2022, placing 19th globally[46]. The government of the Philippines is currently implementing the Gender Equality and Women's Empowerment Plan 2019-2025, which has detailed objectives and measurable targets in the domains of social, economic, and

---

[45] Government of the Philippines. (2021). Philippines National AI Roadmap Retrieved from https://innovate.dti.gov.ph/resources/roadmaps/artificial-intelligence/

[46] World Economic Forum (2022). Global Gender Gap Report 2022. Retrieved at: https://www.weforum.org/reports/global-gender-gap-report-2022/

human development[47]. The Philippines has also had a Data Privacy Act since 2012[48], which will inform future AI policy development, notably in terms of data protection risks for vulnerable population and appropriate data governance.

The Philippines government has focused on the well-being of several groups that intersect with gender rights, namely domestic workers, victims of domestic abuse and migrants, including trafficked persons. For example, the Philippines adopted the 2012 Domestic Workers' Act, which protects domestic workers against abuse. In 2004, it also criminalised domestic violence, and in 2003 instituted policies and practices to eliminate trafficking of persons, most of which are women and girls.

In the digital sphere, the most important documented gaps are in women's participation in the science and technology sphere. Study respondents also described concerns around the use of AI algorithms in human resources and loans, echoing similar concerns globally. Should women or other groups suffer discriminatory treatment from those AI systems, for example by being eliminated from job recruitments or having access to lower loans, this could increase the marginalisation of already vulnerable groups. In addition, certain respondents were familiar with the feminisation of AI, and the risk that feminised interfaces in chatbots or virtual assistants could potentially have an adverse effect on perceptions of women and other groups.

Philippines: developing comprehensive gender data

Low quality of datasets at the expense of women and marginalised groups make it hard to assess the economic, political and social effects of AI technologies. In general, the Philippines lacks comprehensive gender data. Many human security, economic, and environmental indicators are unavailable or lack sex-disaggregated data in both national and international databases.

While the Gender Equality and Women's Empowerment Plan aims to increase economic opportunities for women, several gendered indicators are not available in the Philippines'

---

[47] Open Data Watch. (n.d.). Bridging the gap - open data watch. Retrieved from https://opendatawatch.com/wp-content/uploads/2021/Publications/Bridging-the-Gap-Gender-Data-Asia-Pacific-Technical-Report.pdf

[48] Government of the Philippines. (2012). Republic Act 10173 - Data Privacy Act of 2012. https://www.privacy.gov.ph/data-privacy-act/

national and more so in subnational databases[49]. When gender is not accounted for in AI development, however, AI poses a risk to gender equality. This is even true of initiatives designed to provide an overall social benefit like the use of finance platforms[50] in lending for farmers and expediting the distribution of DSWD's Social Amelioration Funds to beneficiaries. Thus, in addition to high-quality datasets, there is a need for implementable bias mitigation strategies relevant to social benefit systems and financial institutions. Assessing "fairness" in credit scoring is needed such as the likelihood of approval, loan terms, and repayment rate based on gender and social indicators.

Philippines: increasing representation of women STEM fields

Low representation of women in the science and technology labor force leads to various challenges in the design of AI. In the Philippines, as in other countries, there are more men than women graduating from engineering and technology studies, with women accounting for no more than 30 percent of graduates in any academic year during the past 10 years[51].

Similarly, male science and technology graduates have a high labor force participation rate of around 93 percent while only 73 percent of female science and technology graduates entered the job force in 2015[52]. In the same research, there is a fall in female participation in the science and technology labor force after the age of 25, when women begin to marry, and another drop at the age of 35, particularly in the Computing/IT industry. It should also be noted that female S&T graduates' labor force involvement declines until they reach the age of 55[53].

---

[49] Open Data Watch. (n.d.). Bridging the gap - open data watch. Open Data Warch. Retrieved June 10, 2022, from https://opendatawatch.com/wp-content/uploads/2021/Publications/Bridging-the-Gap-Gender-Data-Asia-Pacific-Technical-Report.pdf

[50] DTI. (2021, May 13). Artificial Intelligence Roadmap. Retrieved June 10, 2022, from https://innovate.dti.gov.ph/resources/roadmaps/artificial-intelligence/

[51] Illo, J. (n.d.). Gender profile of the TVET sector - TESDA. Retrieved June 10, 2022, from https://www.tesda.gov.ph/Uploads/File/GAD/2019/Gender%20Profile%20of%20the%20TVET%20Sector%20(final).pdf

[52] Albert, J. R. G Et al. (2020). Future S&T Human Resource Requirements in the Philippines: A Labor Market Analysis. Retrieved from https://kohalibrary.pids.gov.ph/cgi-bin/koha/opac-detail.pl?biblionumber=9767&shelfbrowse_itemnumber=6438

[53] David, C. C., Albert, J., & Vizmanos, J. F. V. (2018). Sustainable development goal 5: How does the philippines fare on gender equality?: Semantic scholar. https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1745_rev.pdf

Lower representation of women in the coding and design of AI and machine learning technologies which can lead to a variety of problems, including the replication of stereotypes, such as the submissive role of voice-powered virtual assistants, overwhelmingly represented by women among others[54].

Philippines: more political empowerment of women

Representation of women in AI is linked to representation in policymaking, as both are areas of influence in today's society. In the Philippines, for example, just 49.1% of women are employed, and those who are employed face salary and income imbalance, with 22 percent of the wage gap and 31 percent of the income gap remaining to be closed. In fact, investing in women's participation in AI can contributing to closing leadership and income gaps, as AI salaries are quite high, compared to other types of work. AI is also one of the main current policy challenges globally, and increasing female voices in this domain can have repercussions for policy leadership more broadly.

---

[54] World Wide Web Foundation. (n.d.). Policy brief W20 Argentina Artificial intelligence. http://webfoundation.org/docs/2018/06/AI-Gender.pdf.

CASE STUDY #1: AI AND HUMAN RESOURCES IN THE PHILIPPINES
BACKGROUND

There has been much uptake globally of AI in job recruitment, in order to increase the efficiency of the process. Instead of reviewing hundreds of resumes manually, AI systems aim to filter out candidates automatically for review. This innovation also addresses a new phenomenon of pandemic resignation, which has led recruiters to spend more time hiring new staff. The Great Resignation, a term coined in 2021 by Anthony Klotz, describes the global trend that many people quit their job during the pandemic[55]. Reasons might vary—health issues, salary consideration, hostile work environments, inflexible work-from-home policies, looking for work-life balance, creating new job opportunities in the hybrid setting, and so on. This phenomenon is leading employers to face labor shortages and fight to recruit top talent for their company. This trend has also been observed in the Philippines, where 62% of HR professionals observed increased resignations in the region during the pandemic[56].

AI RESPONSES

Companies in the Philippines are now using and offering several recruitment platforms that alleviate recruiters' time. For example, MetroCity AI, the company which is part of the Batch 10 startups of UP Diliman's UPSCALE Incubation Program granted by UP Engineering Research and Development Foundation, Inc. and UPSCALE Innovation Hub, developed a recruitment platform that includes a video interview and a CV screening. The platform invites applicants to record a self-introduction video asynchronously by answering questions related to their cultural fit, skills, and behavioural traits for the company. Then, the platform processed applicants' videos if their answers fit well with the company's vision. The platform also filters out unqualified CVs before inviting applicants to the asynchronous video recording[57].

---

[55] Washington Post Live. (September 24, 2021). Transcript: The Great Resignation with Molly M. Anderson, Anthony C. Klotz, PhD & Elain Welteroth. The Washington Post. Retrieved at https://www.washingtonpost.com/washington-post-live/2021/09/24/transcript-great-resignation-with-molly-m-anderson-anthony-c-klotz-phd-elaine-welteroth/

[56] Rivas, R. (June 23, 2022). Great resignation 'in full swing' as Filipinos find better work online. Rappler. Retrieved at https://www.rappler.com/business/great-resignation-filipinos-find-opportunities-hybrid-work-pandemic-sprout-solutions-report-2022

[57] Business World. (May 10, 2022). AI platform to solve the pain of recruitment manual work launched. Retrieved at https://www.bworldonline.com/spotlight/2022/05/10/447421/ai-platform-to-solve-the-pain-of-recruitment-manual-work-launched-2/

CHALLENGES

In the global context, AI recruitment systems have been criticised due to algorithmic bias toward women and marginalised groups. As we have seen, in 2018 Amazon faced important backlash in their gender-biased AI recruitment system. However, concerns remain to this day. As algorithms are trained by patterns of resumes submitted to the company in the last 10-year period, the male-dominant population of the tech industry is reflected in the AI recruitment system. In the Philippines' case, the platform might inadvertently further gender discrimination—down-scoring women of child-bearing age, of different races using different languages, and other marginalised groups.

WHY IT MATTERS

Gender bias in recruitment systems can accelerate gender imbalance in technical or other highly skilled work. If the AI platform continuously excludes women in job positions, this might increase socioeconomic gaps. When the AI recruitment platform is needed, obligatory gender-impact assessment and certification should be required before its deployment for use.

CASE STUDY #2: AI IN MIGRATION MANAGEMENT IN THE PHILIPPINES

BACKGROUND

There are a number of reasons for migration, whether it be voluntary (pursuing economic and academic opportunities), or forced (natural disasters or escaping conflict). The 2022 IOM World Migration Report reported that the number of international migrants has significantly increased in the last decades, to almost 281 million globally[58]. According to UNHCR, 79.5 million people migrated out of necessity at the end of 2019, and 26 million of those people were refugees[59]. In the Philippines, economic migration is very

---

[58] International Orgnization for Migration (2022). World migration report 2022. Geneva, Switzerland: International Orgnaization for Migration.

[59] UNHCR (June 16, 2022). Figures at a glance: 89.3 million people worldwide were forcibly displaced. Retrieved at https://www.unhcr.org/en-us/figures-at-a-glance.html

common[60], with an estimated 10 million Filipinos living abroad and 1 million leaving the country each year[61]. As governments and international organisations seek to manage complex patterns of migration, many have been turning to AI to speed up visa processing times and support migrant tracking. Increasingly, AI technologies have been integrated into the global immigration and migration management.

AI RESPONSES

In order to increase tracking of migrants, case processing and improve communications, AI and other technologies are used increasingly by governments[62]. For example, forecasting tools are used to predict migration population in the Mediterranean, AI-enabled lie detectors and facial recognition systems are used to the European border, and automated decision-making in immigration and refugees application are launched in Canada. These international uses of AI concern Filipinos, who can be subjected to these technologies. The national government may also chose to adopt these technologies itself.

CHALLENGES

While AI tools may increase efficiency in the migration process, supporting migrants in terms of their identification and documentation, and helping governments to track and mange migration flows, there are certain gender and societal risks to consider. For example, biometric identification systems have been shown to be particularly prone to error with women of colour, which can lead to repatriation or other adverse outcomes. Data protection is also critical in the process of migration, as this data could be used to target vulnerable migrants for trafficking, or otherwise manipulate and surveil migrants[63].

---

[60] Maruja M. B. A. (July 12, 2017). The Philippines: Beyond labor migration, toward development and (possibly) return. Migration Information Source, Migration Policy Institute. Retrieved at https://www.migrationpolicy.org/article/philippines-beyond-labor-migration-toward-development-and-possibly-return

[61] International Labour Organisation (2022). Labour migration in the Philippines. https://www.ilo.org/manila/areasofwork/labour-migration/lang--en/index.htm

[62] Molnar, P, (2020). 'AI and Migration Management', in Markus D. Dubber, Frank Pasquale, and Sunit Das (eds), The Oxford Handbook of Ethics of AI (2020; online edn, Oxford Academic, 9 July 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.49

[63] Fournier-Tombs, E. C., Castets-Renard, . (2022). Protecting migrants against the risks of artificial intelligence. IOM Research Series, The impact of COVID-19 on migration and migrants from a gender perspective.

WHY IT MATTERS

Migrants are often in particularly precarious situations, whether they are moving temporarily or more permanently. They are therefore less able than others to appropriately advocate for their rights, particularly when facing discrimination and stereotyping. Additionally, AI systems managing migration are also obscure, making it even more difficult for migrants to identify whether or not an AI system was biased. Much of this challenge relates to divergent intentions on the part of different governments, supporting migrants but also wishing to surveil them. In a Philippines context, given the large migrant workforce, which is notably female dominated in domestic work, it is important to include migrants in those most at risk for AI discrimination.

AI POLICY AND GENDER IN INDONESIA

AI POLICY IN INDONESIA

The Indonesian government has realized the importance of the government's role in implementing[64] AI, with the publication of the book "Indonesia National Strategy for Artificial Intelligence", which is currently in the process of being ratified in a Presidential Regulation on the Acceleration of Implementation of Artificial Intelligence. The draft regulation is in the form of a roadmap that aims to build an ecosystem so that the implementation of AI in Indonesia runs well (optimizing the benefits and mitigating the negative impacts that arise).

The Indonesian National Strategy for Artificial Intelligence Framework was built to support the ideals as Indonesia's Vision 2045, which includes 4 pillars, namely:

    a.   Policy, intended to establish policies and ethics for creating trustworthy AI

    b.   AI Talents, it is intended to give birth to railroad talent along with educational infrastructure and the training.

    c.   Data and Infrastructure, intended to build infrastructure and data ecosystem to support the implementation of the AI.

    d.   Research and Innovation, intended to build AI research results as well as innovative collaboration.

In addition, the roadmap has also set 5 priority areas as an example of accelerating the implementation of AI, namely:

    a.   Health

    b.   Government services and administration

    c.   Education

    d.   Food security

---

[64] The term "Implementation" covers all aspects from the development, use, to the impact caused by AI

　　　e.　　Smart Cities mobility

The Indonesia National Strategy Framework for AI can be seen below:

Figure 4: Indonesia AI Policy Framework
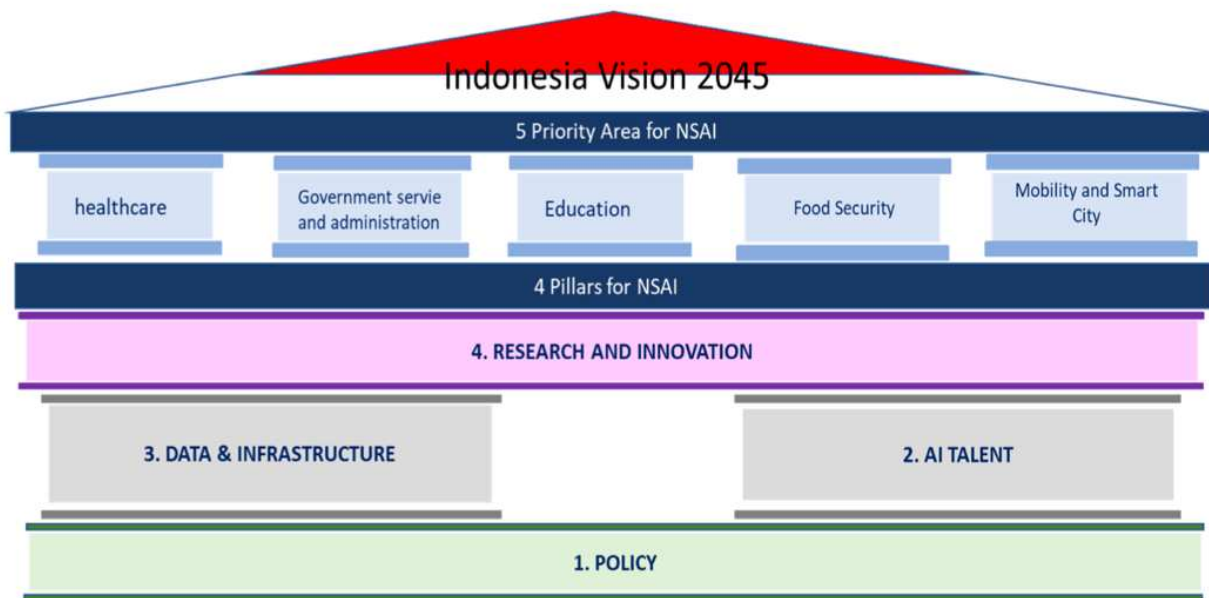
Sumber : BRIN (2022)

NATIONAL PROGRAM PRIORITIES FROM LONG TERM PLANNING DEVELOPMENT 2022-2024

The Government of Indonesia aims to provide AI implementation initiatives that are in line with priority programs that have been scheduled in National Mid-Term Development Plan (RPJMN) 2022-2024[65], especially for long-term initiative programs. Focus areas include 1) strengthening economic resilience for growth; 2) developing areas to reduce gaps; 3) improving quality human resources and competitiveness; 4) mental health and cultural development; 5) strengthening infrastructure to support economic development; 6) building the environment, improving disaster resilience and climate change; 7)

---

[65] Ministry of National Development Planning of Republic of Indonesia. National Mid Term Development Plan 2022-2024.

strengthen stability for POLHUKHANKAM (Politics, Law, Resilience and Security) and public service transformation.

As the largest economy in Southeast Asia, Indonesia has a vibrant startup scene, with companies in the financial technology (fintech) and health tech sectors already established. With a young, internet-savvy population that is increasingly reliant on digital services to meet their requirements, Indonesia is currently a rich environment for startups[66]. Even the tight restrictions opened up new possibilities for enterprising businesses, and platforms for education technology (edtech) started to draw more students who wanted to finish their course or pick up new skills.

Indonesia is currently ranked 79th in the Global Gender Gap Index. As in other countries in the region, gaps are documented in relation to access to the digital economy, along with participation in the AI sector. In addition, this digital divide is greater in rural areas, especially for women, the underprivileged, the elderly, those with lower levels of education and those with disabilities[67].

GENDER RISKS OF AI IN INDONESIA

Several risks were described by study respondents which could be particularly relevant in initiatives by the government to further regulate AI technologies, particularly as they relate to women and other groups.

Certain concerns revolve around the exclusion of marginalised populations from government services using AI. For example, biometric identification systems used by the government to centralise identification and access to services could have exclusionary effect on the elderly, indigenous populations or those living in remote rural areas, due to physical barriers (wrinkled fingers), lack of mobile phones, or language skills. These concerns have also been raised in other uses of biometric identification systems, namely in India[68]. It should be noted, however, that government participants in this study noted

---

[66] Ministry of National Development Planning of Republic of Indonesia. National Mid Term Development Plan 2022-2024.

[67] Nawir, D., & Zultan, A. (2018). Pemberdayaan Masyarakat Wilayah 3T (Terdepan, Terluar, Tertinggal) di Kecamatan Sebatik Barat Kabupaten Nunukan Melalui Revolusi Mental dalam Mewujudkan Indonesia Bersih dan Indonesia Bersatu. Jurnal Pengabdian Masyarakat Borneo, 2(1), 44-50.

[68] Keenan, T. P., & FCIPS, I. (2015). Hidden risks of biometric identifiers and how to avoid them. BlackHat USA.

efforts to provide alternatives who faced barriers to biometric identification in Indonesia, including training and non-AI solutions.

Additionally, as private companies take an increasingly important role in the development of technologies for the public sector, managing sensitive and personally identifiable data, accountability and data protection becomes even more important. For example, data analytics and artificial intelligence tools are used in healthcare, from patient data management, to diagnosis and health communications. In the case of vulnerable populations, any improper data governance or cybersecurity leak could lead to harassment, cyberbullying, stigma, or other socioeconomic harms. In this sense, the Government's current work towards the Personal Data Protection (PDP) law[69], which covers health data, will be particularly relevant.

As in the Philippines, respondents also were aware of risks related to AI for recruitment purposes. However, as one respondent pointed out, prejudice against women and other minority groups also regularly occur with human recruiters. Efforts to reduce bias in AI for recruitment can therefore also be coupled with any effort to reduce hiring bias more generally.

Indonesia and Thailand: reducing the socioeconomic exclusion of women

There remain gaps in Indonesia with respect to gender equality and women's empowerment. The overwhelming majority of households (85%) are headed by men. According to study respondents, as women's work is normally associated with the domestic realm, the interests of women may still not be discussed or addressed directly.

Women in Indonesia make up 56% of graduates across all degrees, and 35% of technology degrees. However, there is still a gap between educational attainment and labour force participation. In Indonesia, women make up 22 percent of the technology workforce. In 2014, approximately 17 percent of the seats in the Lower House of Representatives and 26% in the Upper House, were occupied by women, both figures representing substantial increases in women's representation at the national level. Between 2014 and 2019, 8 out of 34 ministries were led by women[70].

---

[69] Government of Indonesia (2022). Draft Personal Data Protection Law. Available in Indonesian. https://www.dpr.go.id/dokakd/dokumen/K1-RJ-20220920-123712-3183.pdf

[70] FAO (2019).

Indonesia: reduction of accessibility gaps

In Indonesia, the Internet and its supporting infrastructure have not been evenly distributed in rural areas, particularly the disadvantaged, frontier, and outermost (3T) regions. Furthermore, even when Internet infrastructure is available, it does not necessarily ensure that the access to the internet and its use are equal, particularly for women, the poor, the elderly, people with low education attainment and people with disability. And the thirdly, the digital transformation in business expansion has not been fully achieved.

Figure 6: Percentage of Indonesian Households with Internet Access by Province, 2020
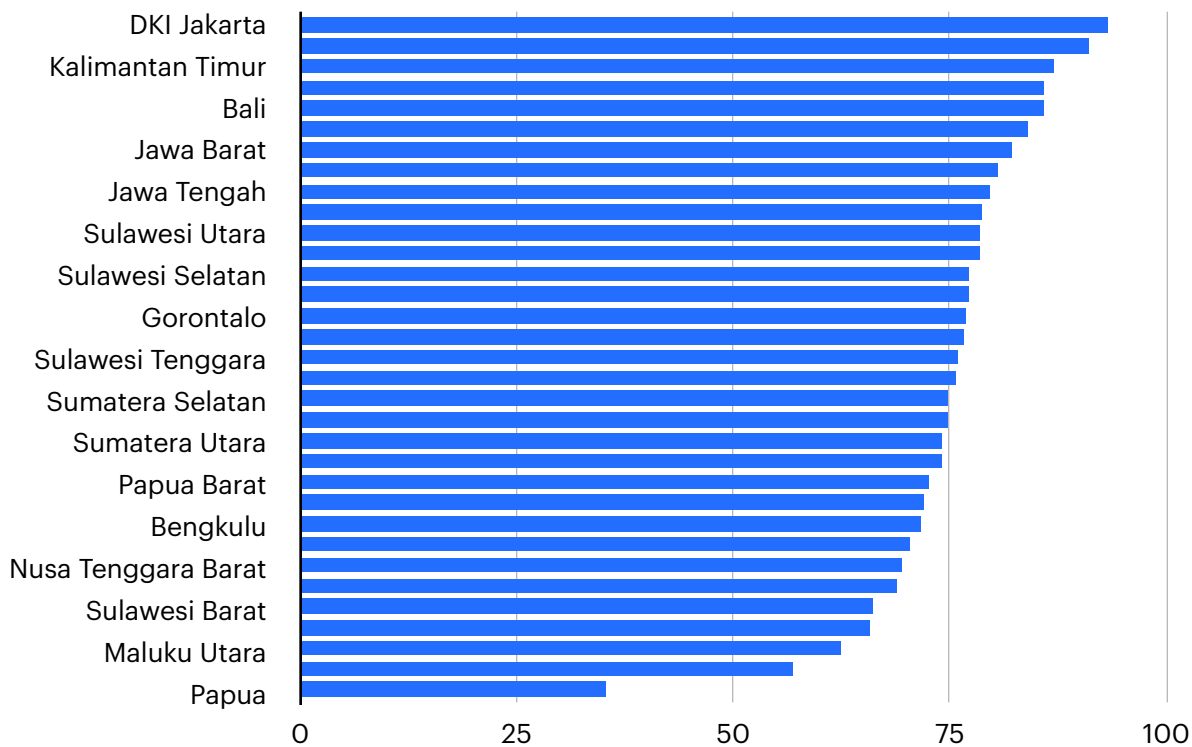
Source: Statistics Indonesia

The figure above shows that the highest ownership of internet access in households in 2020 was in DKI Jakarta Province (93.24 %) and Riau island (91.15%). Papua Province and NTT have the lowest percentage in 2020 with 35.25 % and Nusa Tengara Timur with 56.87%.

Indonesia: Reduction of regional discrimination

Communities that are further away from the provincial capital tend to have uneven infrastructure development. The distance from the capital has caused limited access to information and technology. However, these areas have strong connections to local culture and tradition. The Ministry of Villages, Disadvantaged Regions, and Transmigration (Kemendes PDTT) is formulating new policies to accelerate the development of 62 underdeveloped regions in the next five years. The disadvantaged regions are listed in Presidential Regulation (Perpres) No.63 / 2020 on Determination of Underdeveloped Regions for 2020-2024. In the Perpres, the disadvantaged regions are located in 11 provinces; Papua (21 districts), East Nusa Tenggara (NTT) (13 districts), West Papua (8 districts), and Maluku (6 districts). In 2020, the Ministry of Communications and Informatics played a role as the orchestrator for public communication across ministries and Institutions.

Indonesia: protection of vulnerable children and youth

Children who do not enjoy conventional family life or who are not in school may be vulnerable in digital and AI innovation as well. In today's society, however, use of the Internet among the youth is very high. This, however, leads to an increase in the importance of the digital divide. According to one survey, approximately 92% of youth in urban areas used the Internet over a period of three months, while in rural areas it was only 76%. The data can be used to illustrate the unequal distribution of infrastructure development in Indonesia and the availability of Internet connection services in every area. The Internet use gap also occurs in youth persons with disabilities compared to non-disabilities (41.58% versus 85.99%).

CASE STUDY #1: DRIVER VERIFICATION AND RIDE HAILING APPS
BACKGROUND

In Southeast Asia, Gojek and Grab are the two biggest ride-hailing services[71]. Unlike the traditional taxi business, ride-hailing platforms match passengers and local drivers using their personal vehicles. The platforms build on the digital infrastructures in the regions to offer immediate matching services through passengers and drivers' mobile phone applications. Google and Temasek Holdings reported that ride-hailing services became the largest and fastest growing market in Indonesia[72]. Ride-hailing platforms contribute to resolving Indonesia's traffic as they provide more efficient transportation options for passengers by allowing them to choose either two–wheel rides or four-wheel rides and evaluate drivers' services after using the service. Gojek and Grab have expanded their services for transportation, online food delivery, and online financial service[73].

However, ride-hailing apps run the risk of triggering sexual harassment and abuse as well. Since 2018, many female victims shared testimonies of sexual harassment that occurred on and through ride-hailing apps. For example, local media reported in 2019 that a woman was kidnapped by a Grab Bike driver who drove her to unknown locations[74]. Several other female passengers reported inappropriate physical contact by ride-hailing drivers[75], even when the names of the drivers were known[76]. As sexual harassments on ride-hailing apps increased, Indonesian women united on social media campaigns advocating for more concerted action towards the protection of female passengers.

---

[71] Sheng, Charlie. (January 18, 2022). Ride-hailing race in Indonesia: Gojek versus Grab. Measurable.AI.

[72] The ASEAN Post (2018), "2018: A transformational year for ride hailing", The ASEAN Post, available at: https://theaseanpost.com/article/2018-transformational-year-ride-hailing (accessed 23 July 2020).

[73] Almunawar, M.N., Anshari, M. and Ariff Lim, S. (2021), "Customer acceptance of ride-hailing in Indonesia", Journal of Science and Technology Policy Management, Vol. 12 No. 3, pp. 443-462. https://doi.org/10.1108/JSTPM-09-2019-0082

[74] Dwi Bowo Raharjo & Achmad Fauzi. (August 13, 2019). Driver Ojol Lecehkan Penumpang, Grab Diminta Buat SOP Cegah Pelecehan. Suara.com. Retrieved at https://www.suara.com/news/2019/08/13/202218/driver-ojol-lecehkan-penumpang-grab-diminta-buat-sop-cegah-pelecehan

[75] A video uploaded by @dearcatcallers.id: https://www.instagram.com/p/B0ikXToAJiQ/

[76] Mei Amelia R. (Feb 13, 2018). Driver Lecehkan Penumpang, Ini Tanggapan Manajemen Go-Jek. Detiknews. Retrieved at https://news.detik.com/berita/d-3865895/driver-lecehkan-penumpang-ini-tanggapan-manajemen-go-jek

AI RESPONSES

The Ministry of Transportation's regulation no. 118 of 2018 reacted to the country's issues by stating drivers' obligations to provide appropriate protection to their passengers. Ride-hailing apps agreed to follow the legal aid and sought the help of the Indonesian Women's Commission. They also updated preventive measures using digital technologies such as a Share My Ride feature, selfie authentication, surveillance camera, and emergency buttons. The Indonesian ride-hailing apps integrated facial recognition technology into its mobile application. Gojek drivers are now required to provide a selfie when they log into the account and accept a call[77]. Facial recognition AI then verifies drivers' identities before accepting the call.

OPPORTUNITIES

When the verification system is successfully used, facial recognition systems can contribute to not only preventing in-app fraudulent activities but also making women and other marginalised clients less vulnerable to kidnapping and sexual violence. By knowing drivers' identities in advance, women and marginalised populations can have more agency in using ride-hailing apps for their convenience.

WHY IT MATTERS

AI technology, such as driver verification and ride hailing apps, has shown that AI technology is able to support gender sensitivity, so that it can be beneficial for women to be safer and more protected, which of course will further realize the fulfillment of women's rights. Until now related to this application, there have been no incidents that show differences in treatment by this technology which is gender biased. Further attention and research are needed regarding this application (and similar applications) in providing fair decisions (especially for gender issues etc.).

---

[77] Rani Rahayu (August 21, 2019). In Indonesia, ride-hailing apps fail to adequately address sexual abuse. Vice. Retrieved at https://www.vice.com/en/article/59npy3/indonesia-ride-hailing-apps-sexual-abuse-grab-gojek

CASE STUDY #2: AI SUPPORT FOR CYBERBULLYING
BACKGROUND

As the digital platforms have expanded, cyberbullying has become common worldwide. About 59% of American teens have experienced some type of cyberbullying online or on their mobile phones[78]. People bully and/or harass others on digital spaces, such as social media sites, gaming platforms, and mobile applications, by threatening messages, posting rumors and embarrassing photos/videos, remarking sexual messages, or imposing hate speech[79]. Cyberbullying has been recognised as a serious social issue in Indonesia as well. Several researchers reported that Indonesia ranked third place in cyberbullying after Japan and South Korea among 40 Asian countries surveyed[80] and more than 91% of Indonesian citizens are aware of the fact that their children were bullied on social media[81].

AI RESPONSES

As cyberbullying becomes a global issue, social media platforms, such as Instagram, developed new tools using machine learning techniques to detect bullying in images and captions[82]. However, monitoring images and captions against cyberbullying in Southeast Asia has been limited due to the language difference and the lack of data set[83]. As another form of digital response, an Indonesian start-up - Bully ID, developed an online platform that provides psychological and legal support to victims and survivors of

[78] Anderson, M. (September 27, 2018). A majority of teens have experienced some form of cyberbullying. Pew Reserach Center. Retrieved at https://www.pewresearch.org/internet/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/

[79] UNICEF (February 2022). Cyberbullying: What is it and how to stop it. Retrieved at https://www.unicef.org/end-violence/how-to-stop-cyberbullying

[80] Kaman, C. (2007): What country has the most bullies?, Latitude News, viewed 29/4/2013 2013, http://www.latitudenews.com/story/what-country-hasthe-most-bullies/

[81] Gottfried, K. (2012): One in Ten (12%) Parents Online, Around the World Say Their Child Has Been Cyberbullied, 24% Say They Know of a Child Who Has Experienced Same in Their Community, Ipsos, viewed 1/4/2013 2013, http://www.ipsoshk.com/wpcontent/uploads/2012/04/Cyberbullying-factumAP.pdf.

[82] Frick Co. & Stroud, S. R. (November 20, 2018). Filtering out cyberbullying. Center for Media Engagement, The University of Texas at Austin. Retrieved at https://mediaengagement.org/research/filtering-out-cyberbullying/

[83] Nurrahmi, H. & Nurjanah, D. (2018). Indonesian Twitter cyberbullying detection using text classification and user credibility. 2018 International Conference on Information and Communications Technology. DOI:10.1109/ICOIACT.2018.8350758

harassment and cyber violence in Indonesia. Bullyid App provides a live chat for victims and survivors who need free psychological and legal guidance. The company also removes unwanted intimate image/videos online for victims of revenge porn/non-consensual intimate image abuse[84].

OPPORTUNITIES

As women are one of the main targets of cyberbullying, the online platform empowers victims to make their next move. The application is supported by licensed psychologists, lawyers, and chatbot systems to help victims confidentially. The counselling program might also help researchers to develop an Indonesia-based monitoring system for cyberbullying with their consent, by gathering information about patterns of cyberbullying behaviours and languages.

WHY IT MATTERS

As case study #1, this technology is also useful in providing protection to women or those who have weak (marginalized) social power. However, further research is needed whether this application has included data about cyberbullying that is inherent in all groups. Do not let the application that is developed only defends bullying in a group, while other groups are not facilitated.

---

[84] Pasaribu, A. (Septermber 20, 2022). Bullying Indonesia. NAMATI. Retrieved at https://namati.org/network/organization/bully-id/

AI POLICY AND GENDER IN THAILAND

AI POLICY IN THAILAND

Thailand has both an AI strategy[85] and an AI ethics framework [86]. The AI strategy is part of the Thailand 4.0 policy, which was launched in 2016 with the objective of shaping the economy for a 20-year period. The Thai government sees AI as a critical component of its strategy to escape the middle-income track and join the ranks of the world's wealthy nations. AI being a cornerstone of Thailand 4.0. According to the Thailand 4.0 policy, AI will be important in many areas some of which include education, industry, agriculture, trade, and business development.

The AI policy was part of the Thailand National Action Plan on Artificial Intelligence for the Development of Thailand (2022-2027). The Cabinet resolution on July 26, 2022 approved this National Action Plan that was proposed by Office of the National Digital Economy and Society Commission (ONDE), Ministry of Digital Economy and Society (MDES) and Ministry of Higher Education, Science, Research and Innovation. On August 17, 2022, the National AI Committee was appointed by the Prime Minister (General Prayut Chan-ocha).

Thailand also created the Guidelines on AI Ethics to serve as a guideline for researchers, designers, developers, and service providers of artificial intelligence (AI) systems. The document has six main areas of focus: (i) competitiveness and sustainable development; (ii) law, ethics and international standards; (iii) transparency and accountability; (iv) security and privacy; (v) fairness; and (vi) reliability. Within these six focus areas, the document establishes the roles and responsibilities of three main stakeholder groups - government agencies and regulators; researchers, designers, developers, and service providers; and users.

THAILAND AI ETHICS GUIDELINES

The cabinet resolution on February 2, 2021 approved this guideline proposed by the Office of the National Digital Economy and Society Commission (ONDE), Ministry of Digital Economy and Society (MDES). The Thailand AI Ethic Guidelines consist of ethical

---

[85] Government of Thailand. (2016). Thailand 4.0 Policy.

[86] Government of Thailand. (2020). National AI Ethics Guideline. Retrieved from https://onde.go.th/assets/portals/1/files/Thailand%20AI%20Ethics%20Guideline%20(White%20paper)%20Edit%20Version.pdf

principles. ONDE is taking leadership for the guidelines as a pilot in the first phase and evaluating the results of the operation before expanding the results in the next phase[87].

The Guidelines on AI Ethics begins with a chapter that provides an exhaustive analysis of current standards from across the globe, including the European Council, The United States, Dubai, Singapore, UNESCO, Microsoft, and China. Then it presents the six major "AI Ethics Principles". In the section below, we spend some time expanding on the principles as an important normative document for the region.

1. Competitiveness and Sustainability Development: creation and use of AI systems that benefit people and the environment while also contributing to human competitiveness and fair economic development. AI systems should be continuously developed in order to contribute to further innovations and the development of innovative industries.

2. Laws, Ethics, and International Standards: AI systems should be established in accordance with laws, norms, ethical principles, and international standards, to respect human rights and dignity. AI system designs should be human-centered with humans making final decisions based on AI system recommendations. Autonomous systems should never be used to decide the fate of a human being.

3. Transparency and Accountability: developers and designers of AI systems should be held accountable for their work; AI algorithms must be traceable and diagnosable; and AI should be built in a transparent manner. Moreover, researchers, developers, designers, and service providers should all be accountable for the effects of using AI.

4. Security and Privacy: there are four principles relating to security and privacy: 1) AI must always be built to serve humans rather than control, resist, or attack them; 2) AI should always be designed to be resistant to external threats and to preserve the integrity of its data; 3) AI should have mechanisms for human intervention in the system in order to control any risks that may have impact on humans; 4) national authorities must collaborate with international agencies to prevent the emergence of autonomous weapon systems.

---

[87] Office of the National Digital Economy (2022). ONDE AI Ethics. Retrieved from: https://ethics.tu-onde.com/

5. Fairness: AI technologies should be developed with concern for diversity, should not entrench monopolies, should improve fairness and justice, and should empower people who are underprivileged in society. In this section, the guideline does make mention of bias in datasets, which could serve as a basis for further exploration of gender-based risks.

6. Reliability: Trust in AI systems is essential. AI systems must therefore be sufficiently accurate in their predictions and produce results that are reliable and reproducible. Moreover, AI systems must utilise quality control and data integrity systems involving feedback from users[88].

These principles are to be implemented differently by the three main stakeholder groups - government agencies and regulators; researchers, designers, developers and service providers; and users. Consideration of private sector interests has been part of the guideline development in Thailand.

In addition to the AI Ethics Principles and Guidelines, the National Science and Technology Development Agency published an AI Ethics Guideline in March 2022[89], which is to be used by the agency internally and with all grant recipients and contractors. This guideline refers to seven principles of AI ethics, namely privacy, security and safety, reliability, safety and non-discrimination, transparency and explainability, accountability, and human control for sustainability of humanity. The document makes certain references to gender biases, notably in the inclusion of sexual orientation as sensitive personal data; as well as the reference to Buolamwini and Gebru's work on gender and racial discrimination in facial recognition[90], which was also discussed earlier in this report.

---

88 Government of Thailand (2019). AI Ethics Principles and Guidelines. Retrieved at: https://www.etda.or.th/getattachment/9d370f25-f37a-4b7c-b661-48d2d730651d/Digital-Thailand-AI-Ethics-Principle-and-Guideline.pdf.aspx?lang=th-TH

89 NSTDA (2022). NSTDA AI Ethics Guideline. Retrieved from: https://waa.inter.nstda.or.th/stks/pub/ori/docs/20220831-aw-book-ai-ethics-guideline.pdf

90 Buolamwini, J. & Gebru, T. (2018).

Two other documents on AI ethics published by the Thailand Digital Government Development Agency also refer to bias - AI for Government Administration and Service[91], which refers to Amazon's gender-biased recruiting tool; and the AI Government Framework[92], which highlights the importance of AI safety but does not refer to gender risks explicitly.

GENDER RISKS OF AI IN THAILAND

The Guidelines on AI ethics does document a lower participation in the digital sphere for women and other marginalised groups, including lower participation in STEM education and professional fields. They are also subject to the digital divide, being less likely to have access to the Internet or Internet-connecting devices.

Thailand: stereotyping and AI datasets

There exist important gender-based considerations in Thailand that could be integrated in the future. Thailand is 79th globally in the 2022 Gender Gap Index, with a slight decrease in score from the previous year. As in all cultures, stereotyping is present in Thai culture, with the presentation of women in domestic roles as the most common stereotype. While there is a stronger LGBTQ+ presence in popular culture than other countries in the region, individuals may still be discriminated against or misrepresented, which may translate into biases in AI training data. The same goes for ethnic minority groups, which might be subjected to discrimination online.

In addition, respondents highlighted several concerns in relation to AI use in Thailand, which, while they may be covered by the ethics guidelines, may warrant further development in a gender module.

First, there was concern as to the lack of representation of women, LGBTQ+ persons and other community members in the datasets, as well as testing and monitoring protocols. As we have seen, this can take place both because data is aggregated and gender differences might be lost, and because gender impact assessment does not take place.

---

91 DGA (2022). AI for Government Administration and Service. Retrieved at: https://www.dga.or.th/document-sharing/dga-e-book/79539/

92 DGA (2021). AI Government Framework. Retrieved at: https://dgti.dga.or.th/e-books-ai-government-framework/

Thailand: protecting migrants

As in the Philippines, Thailand is also a country with significant migration flows, notably for labour. Concerns were raised about the manipulation of social media recommendation systems to recruit unwitting migrants for human trafficking, especially in relation to sexual exploitation.

Additionally, as we will see in the case study, exclusion of women and marginalised groups in society took on special meaning in AI, as an investment in smart farming tools may have uneven impacts. Respondents suggested efforts to ensure that the benefits of new AI investments were spread evenly throughout different communities.

Finally, the research demonstrated low awareness of gender risks in AI, particularly in the private sector. This could point to an opportunity in Thailand, and in other countries, to conduct gender sensitivity training to AI developers and investors in the private sector.

CASE STUDY #1: AI CHATBOTS FOR SURVIVORS OF VIOLENCE
BACKGROUND

According to World Health Organisation (WHO), one in three women is affected by violence and the threat of violence globally[93]. In Thailand, from 2008 to 2013, there were 30,000 cases of rapes occurring each year and 4,000 cases reported to the police per year[94]. However, reported cases of sexual violence are limited in that, on average, only 11 percent of women who have experienced sexual assault have reported it to the police[95]. According to the Promotion of Family Institution Development and Protection Act[96], in Thailand, committing an act of domestic violence is imposed up to six months imprisonment and/or 60,000 baht fine. Additional settlement money might be charged to the survivor. Despite the legal support for domestic violence, many women have limited knowledge and limited access to information about what support they could get and how they report the crime[97].

AI RESPONSES

To help survivors, a Thai policy investigator, Lieutenant Colonel Mekhiyanont, developed a chatbot, the Sis Bot, which is accessible from a mobile device or a computer. In the case of an incident, women facing domestic violence can message the Sis Bot which will respond 24/7 with information about how to report the incident to the police, how to preserve evidence, and what support services or compensation they are entitled to by law. The chatbot is also used as a learning tool to understand a step-by-step guide on

---

[93] World Health Organization, Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-partner Sexual violence (2013), p.2. Available from: http://apps.who.int/iris/bitstream/10665/85239/1/9789241564625_eng.pdf

[94] Thailand Development Research Institute (25 July 2014) Factsheet on Rape: Some Observations. https://tdri.or.th/2014/07/tdri-factsheet-24/

[95] UN Women, Progress of the World's Women: In Pursuit of Justice. (New York, 2013).

[96] National Legislative Bodies (25 July 2007). Thailand: Act No. B. E. 2562, on the Family Protection and Development. Available on https://www.refworld.org/docid/4a5460692.html

[97] UN Women, UNODC and UNDP (2017). The trial of rape: Understanding the criminal justice system response to sexual violence in Thailand and Viet Nam. available at https://asiapacific.unwomen.org/en/digital-library/publications/2017/09/the-trial-of-rape#view

how to support survivors in the community and what advice society should give to survivors of gender-based violence[98].

OPPORTUNITIES

In many cases, women and girls who wish to seek help do not know whom they should consult and how they start the process. Also, the experiences of sexual violence might stigmatise the victims, making survivors stop seeking support due to the feeling of shame. By communicating with the Sis Bot, the survivors can be aware of their rights and the possible actions to take. Even if the survivors do not want to open a police investigation, they can receive other options, such as shelters, counselling services, and other social supports for their experiences. The AI-based chatbot can help women take the first step to seek the information they need without having a fear of social stigma and judgment.

WHY IT MATTERS

This is another example of a use of AI that can protect women and enhance gender rights, rather than the other way around. In this regards, Sis Bot provides AI-based support for gender-based violence and also empowers women to make informed decisions. It can equip survivors with all the facts about the legal process and other options, which provide the rights of access to justice for women. Policy makers can therefore invest in technologies of gender-responsive justice systems promoting human rights, gender equality, and effective participation.

CASE STUDY #2: SMART FARMING TOOLS IN THAILAND
BACKGROUND

According to Thailand's National Statistics office, more than 40% of Thailand's workers are employed in agriculture as their core economic activity. Despite the large portion of workers in the sector, agriculture contributes only 8.55% of GDP in 2021[99]. According to Deputy Prime Minister, with the ageing of the population, Thailand faces fewer workers

---

[98] UN Women (30 May, 2019). Using AI in accessing justice for survivors of violence. Avaaible at https://www.unwomen.org/en/news/stories/2019/5/feature-using-ai-in-accessing-justice-for-survivors-of-violence

[99] O'Neill, A. (Sep 9, 2022). Share of economic sectors in the GDP in Thailand 2021. Statista. Retrieved at https://www.statista.com/statistics/331893/share-of-economic-sectors-in-the-gdp-in-thailand

from the younger populations in the agricultural sector[100]. As global societies become urbanised, Thailand's government has implemented national programs for agricultural sectors to meet the global demand for agricultural products and support almost one-third of the country's labour force in this sector. For example, Thailand implemented the young smart farmer programme as the national development plan for agriculture. The government also launched the "Thailand 4.0" economic model, which focuses on technological developments in agricultural production and its commercialisation. Before Thailand, Hokkaido, the northernmost of Japan's main islands, successfully launched agricultural technology and ecotourism by attracting 235,000 Thai visitors in 2018[101]. Thailand's government decided to apply Hokkaido's smart farming tourism tool[102] to upgrade community tourism and to raise the annual incomes of farmers from the current 57,450 baht to 390,000 baht within the next 20 years[103]. Thailand expects to attract international visitors through a combination of tourism and farming as well as attracting a younger workforce for advanced technology with increased productivity.

AI RESPONSES

"Thailand 4.0" uses advanced technologies such as biotechnology, robotics, sensors, drones, big data, artificial intelligence, and the Internet of Things (IoT). In its essence, Thailand integrates smart farming tools into three key processes of agriculture: observation, diagnostics, and decision and implementation. Using sensors, farmers can collect data from the farm, which may include soil conditions, temperature, pressure, and humidity, which are then analysed using AI models. AI-powered drones also support data collection through near-real-time evaluation, and robotics automate crop harvesting or other manual labour tasks[104].

---

[100] Sharon, A. (Dec 3, 2019). Thailand pushing smart agriculture. OpenGOV. Retrieved at https://opengovasia.com/thailand-pushing-smart-agriculture/

[101] Temmen, N. & Schilling, J. (2021). Smart farming technology in Japan and opportunities for EU companies. EU-Japan Centre for Industrial Cooperation.

[102] Theparat, C. (Nov 26, 2019). Somkid touts Hokkaido smart agricultural model. Bangkok Post. Retrieved at https://www.bangkokpost.com/tech/1802294/somkid-touts-hokkaido-smart-agricultural-model

[103] Patanapichai, K. (2022). A new wave of smart farmers yielding Thai crops. Thailand Now. Retrieved at https://www.thailandnow.in.th/business-investment/a-new-wave-of-smart-farmers-yielding-thai-crops/

[104] Tech Wire Asia (Nov 13, 2019). What Thailand plans to invest billions in smart farming initiatives in 2020. Retrieved at https://techwireasia.com/2019/11/why-thailand-is-investing-in-smart-farming-in-2020/

CHALLENGES

Even though AI in agriculture is a government priority, access to smart farming tool might not be equal for all workforces. With the requirement of appropriate literacy for advanced technologies, the existing digital divide might remain the same or get worse, which might link to gender, racial, and socioeconomic marginalisation. Additionally, as integrating smart farming tools costs a lot of money, Thailand's government provides farmers with loans at low-interest rates. Given that loans are more challenging to obtain for women and other marginalised groups, including because of decisions made by AI algorithms, it would be important to ensure that these groups can, in fact, benefit from the new tools.

WHY IT MATTERS

Smart farming tools can be extremely beneficial for the national economy, but, without a proper approach for already marginalised populations in the sector, the new AI-based industry might widen the divide between those who do not have access and those who do have it. The national plan does not include a literacy programme or support program for marginalised populations, including women and land tenants. Access to AI innovations such as smart farming tools could be prioritised in a gender-sensitive AI policy, including policy-level support for a proper literacy program, special loans for marginalised populations, and a technological support system.

AI POLICY AND GENDER IN MALAYSIA

AI POLICY IN MALAYSIA

In 2021, the Malaysian Prime Minister launched the Malaysia Digital Economy Blueprint, to be implemented in three phases until 2030[105]. The plan seeks to capitalise on the transformative potential of digital technologies, especially as seen during the global pandemic. This Blueprint complements the 12th Malaysia Plan and the Vision for Shared Prosperity 2030. Malaysia also has a focus on the Fourth Industrial Revolution, with the National 4IR Policy being promulgated by the Economic Planning Unit of the Prime Minister's Office.

THE NATIONAL AI FRAMEWORK OF MALAYSIA

Malaysia has also established the Malaysia Artificial Intelligence Roadmap 2021-2025[106] with 6 strategies and 22 strategic initiatives that aim to augment economic prosperity and social well-being. Malaysia's AI roadmap targets 5 priority areas: agriculture and forestry; medical and healthcare; smart cities and transportation; education; and public service. The national AI roadmap is guided by the Principles of Responsible AI, which emphasises: fairness; reliability, safety and control; privacy and security; inclusiveness; pursuit of human benefit and happiness; accountability; and transparency. This roadmap was developed by the Ministry of Science, Technology and Innovation (MOSTI), assisted by experts from the Universiti Teknologi Malaysia.

The focus of this policy is on ease of doing business while also promoting their own industry and enterprise, with innovation and cross-sectoral integration and adaptability becoming important goals. On the governance front, efficiency is underscored, and modernised administration is outlined as a goal. Data-driven national planning and technology-aided administration are indicated as priorities for the government.

Further, human-centricity is cited as a priority in the proposed technological revolution that the Malaysian government aims to propel, through the National 4IR Policy. The document suggests that citizens should be able to make use of technological advancements, while preserving humanistic and societal values, and cultural heritage. In addition, the whole-of-nation approach is outlined, wherein the idea that no one shall be

---

[105] Government of Malaysia. (2021b). Malaysia Digital Economy Blueprint. Retrieved from https://www.epu.gov.my/sites/default/files/2021-02/malaysia-digital-economy-blueprint.pdf

[106] Government of Malaysia. (2021a). Malaysia Artificial Intelligence Roadmap. Retrieved from https://airmap.my/

left behind is underscored. Onus is also placed on citizens to be more aware and use emerging technologies for good.

GENDER RISKS OF AI IN MALAYSIA

In terms of gender rights, Malaysia has its own concerns, with a 103rd ranking in the Global Gender Gap Index. Research respondents in the civil society, government and academia were familiar with possible gender and societal risks of AI, notably the known examples of discrimination in AI recruitment tools and in loan allocation. However, this familiarity was much lower in the private sector. As in other countries, this points to an opportunity for gender sensitivity training for the developers and investors in AI technologies.

Additionally, concerns were also raised about the use of social media to promote trafficking or to share sexualising content about women, furthering negative stereotypes which could lead to gender-based violence.

Finally, respondents raised the importance of higher representation of women and other marginalised groups in AI and policymaking, leading to reprioritisation of gender considerations in AI development.

A gender-specific agenda in Malaysia could combine several elements, from fostering gender sensitivity in smart city development, to promoting more technologies for women's health and women's issues. As it stands, there are few mechanisms to monitor and protect women and marginalised groups from the harms of AI, if they do occur. While human rights mechanisms can support women after the fact, the obscurity of AI tools means that it is often very difficult to identify whether or not a harm has occurred.

Malaysia: recognition of social changes driven by AI

Data-driven governance is seen as a priority in data and technologies strategies in the country. However, specific ways of ensuring data protection and privacy for marginalised groups can still be further detailed. The national move towards hybrid cloud computing for public data is also important. The labour force in Malaysia is highly productive for the region (Global Competition Report, 2020), but numerous shifts due to AI-driven platformisation have altered work conditions in Malaysia. Gender impact analysis of these

changes would be important, as they also relate to other marginalised groups in the workforce.

Malaysia: development of education and AI literacy

The National 4IR Policy mentions the importance of AI literacy and awareness. Education can therefore be an area where gender-equitable and gender-sensitive education can play an important role, not only for women but also for men.

CASE STUDY #1: AI FOR BREAST CANCER DETECTION
BACKGROUND

There have been many deployments of AI in the healthcare industry globally. With the advent of 21st-century global pandemics such as Ebola and COVID-19, many companies have invested very large funds to develop AI technologies in healthcare. There is global optimist that applying AI can increase the effectiveness of daily medical tasks, diagnostics and treatment[107]. In Malaysia, medicine and healthcare are one of the national priority areas in AI development and regulation. During COVID-19, Malaysia used AI to distribute vaccines, which reduced supply spending by 20% and logistic costs by 40%. There has also been AI-based diagnostic and predictive systems, such as Personalised Proactive Healthcare, Autonomous A-eye System (preventing blindness), and AI-Nasoalveolar System (AI-assisted predictive model[108]).

AI RESPONSES

Given that disease screening is one of the national priorities of AI in healthcare, there is ongoing research on AI for breast cancer prediction and screening[109]. Many clinicians have striven to detect early stage breast cancers and reduce mortality through screening mammography and breast ultrasounds. Artificial intelligence-based screening tools have been introduced with the promise of increasing diagnostic accuracy. With deep learning techniques, these tools can predict patients' clinical symptoms and detect clinically relevant features in imaging data that human eyes cannot perceive[110]. These tools have been adopted in hospitals globally, including in Malaysia.

OPPORTUNITIES

Using AI for female health concerns is a big opportunity, as this field is generally underserved in the healthcare systems. Given that breast cancer is also a growing

---

[107] Shairi, I. S. (August 13, 2021). Malaysia: Robots to the rescue: Regulation of artificial intelligence in healthcare in Malaysia. Azmi&Associates.

[108] Ministry of Science, Technology & Innovation (2021). Malaysia national Artificial Intelligence roadmap 2021-2025.

[109] Ganggayah, M. D., Dhillon, S. K., Islam, T., Kalhor, F., Chiang, T. C., Kalafi, E. Y., & Taib, N. A. (2021). An artificial intelligence-enabled pipeline for medical domain: Malaysian breast cancer survivorship cohort as a case study. Diagnostics, 11(8), [1492]. https://doi.org/10.3390/diagnostics11081492

[110] GE Health Care Malaysia (Sep 06, 2021). How AI is changing the game in breast imaging. GE Health Care Malaysia. Retrieved at https://www.gehealthcare.com/en-my/insights/article/how-ai-is-changing-the-game-in-breast-imaging

concern in Southeast Asia, this fosters regional awareness as to the importance of early detection and prediction for disease prevention.

WHY IT MATTERS

Additionally, while AI has tended to be developed without the participation of women, using AI for female health concern and increase inclusion, as it requires a focus on women in datasets, testing, and monitoring. Gender-sensitive AI can therefore mean not only assessing risks in AI deployments but also developing AI solutions that can promote the well-being of women. There is a need for a policy support promoting these types of AI solutions for women.

CASE STUDY #2: AI AND SMART CITIES

BACKGROUND

Smart cities are the new way of conceptualising urban development in the digital age, using information and communication technologies to booster a city's efficient operation. The definition of smart cities varies, but, according to the World Bank, a smart city uses data and digital technologies in planning and managing the city's core functions to become efficient, innovative, inclusive and resilient[111]. Malaysia has also reflected the global trend by adopting the Malaysia Smart City Framework (MSCF). As the definition of smart cities might differ depending on the national context[112], the Malaysian concept of the smart city means urban planning, development, and management that can promote social/economic growth to solve urban challenges of the region, such as the inefficient delivery of urban services, environmental pollution, and traffic congestions[113].

AI RESPONSES

There are seven major components of smart cities described in the MSCF: 1) smart economy, 2) smart living, 3) smart environment, 4) smart people, 5) smart government, 6)

---

[111] Kahn, M. E. (2014). Sustainable and smart cities (English). Policy Research working paper ; no. WPS 6878 Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/784061468155133227/Sustainable-and-smart-cities

[112] United Nations Development Programme (Sep 15, 2021). Why truly smart cities are crucial for development: How Tehran is exploring technology and innovation to improve urban lives and livelihoods. UNDP Asia and the Pacific. Retrieved at https://www.undp.org/asia-pacific/blog/why-truly-smart-cities-are-crucial-development

[113] Bachok, N. (2022). Malaysia smart city framework. The Government of Malaysia's Official Gateway. Retrieved at https://www.malaysia.gov.my/portal/content/30947

smart mobility, and 7) smart digital infrastructure. Artificial intelligence is a key component of Malaysia's smart city project. One active focus area has been AI-driven public transport, which provides AI-driven autonomous maintenance, repair, and operation of public transport[114]. There is also a possibility of using AI surveillance systems based on facial and object recognition, behavioural and movement analysis algorithms to analyse live security video to reduce potential risks and threats[115].

CHALLENGES

Smart cities involve a complex interplay of AI technologies and data, which are generally not audited for gender and societal impact. Urban planning and the design of technology are political and can intentionally marginalise socioeconomic groups, including gender and race[116]. Without auditing gender needs, urban design and its artefacts might disempower women and other social groups.

WHY IT MATTERS

When developing training datasets for smart cities, data might not accurately reflect women' needs or differences. For example, gendered mobility patterns are not reflected in the design of mobility in smart cities, which is a national focus. As another example, surveillance cameras with facial recognition might help protect the security of smart cities, but women wearing hijabs might not be properly identified by the image recognition system and the overall governance system. In smart city planning, policymakers could use a gendered understanding of AI in the national context. Given the multiple layers of data, AI and IoT technologies used in smart cities, policy frameworks governing them should be tied to broader AI policies in Malaysia and each country in the study.

---

[114] Ministry of Science, Technology & Innovation (2021). Malaysia national Artificial Intelligence roadmap 2021-2025.

[115] Telekom Malaysia Berhad (2022). Demystifying technology: The DNA of a smart city. Retrieved at https://www.tmone.com.my/resources/think-tank/article/ipoh-smart-city/

[116] Winner, L. (1980). Do Artifacts Have Politics? Daedalus, 109(1), 121–136. http://www.jstor.org/stable/20024652

OPPORTUNITIES IN AI POLICY IN THE REGION

In the following section, we provide more specific opportunities in AI policy in the four above countries, notably those that relate to mitigating risk in AI for women and other societal groups. These opportunities are common to the different countries surveyed, although may be implemented in different ways. They can be categorised broadly as follows:

- Development of AI societal impacts committee or task force

- Creation of gender and AI guideline

- AI safety standardisation across the region

- Investments in women in AI networks, training, company development and policy participation

OPPORTUNITY # 1: DEVELOPMENT OF AI SOCIETAL IMPACTS COMMITTEE OR TASK FORCE

As we have seen the societal impacts of AI are complex and far-reaching. While risks can be similar from one country to the next, there are also challenges and needs in each country which need to be addressed specifically. In order to monitor these societal impacts, this report recommends the development of a committee or task force with a clearly identified chair or lead institution, which would lead gender and societal-impact assessment for high-risk uses of AI.

The first responsibility of this committee would be to revisit the high-risk uses of AI outlined in the European Union's Draft AI Act, which have been discussed in this report and are listed below:

Table 2: Mapping of high-risk uses of AI

High-risk use of AI (EU)	Example use	Possible risks
Biometric identification	Use of biometrics for voting, visa processing or other government services	Exclusion of marginalised populations; misidentification (mismatch or other identification error) with severe impacts on marginalised populations
Management and operation of critical infrastructure	Management and optimisation of smart cities (such as healthcare, energy, traffic) using AI systems	Exclusion or misallocation of services for marginalised populations; privacy risks with severe impacts for marginalised populations
Education and training	Tracking progress and identifying at-risk students; recommending training content	Overly generalised approach leading to exclusion or discrimination against students; stereotyping of recommendation system based on group belonging
Employment	Use of AI system to pre-screen resumés for job applications	Automating rejection of resumés based on discrimination (gender or other group identification)
Access to essential services	Service allocation through AI optimisation (at community or even individual level)	Exclusion of marginalised populations or individuals
Policing	Crime prediction for allocation of police resources to high-risk areas	Over policing of high-risk areas, leading to confirmation bias
Migration, asylum and border management	Use of AI system to fast-track decisions on visa applications	Discriminatory decision-making (rejection based on group belonging, rather than actual risk)
Administration of justice and democratic processes	Use of AI systems to propose appropriate sentence length, or calculate risk of recidivism	Discriminator decision-making (longer sentence based on group, rather than actual crime)

While the above risks apply to the focus countries studied here, there may be additional uses of AI considered by the commission that were not considered in the European Union. This can include some of the uses of AI discussed in the case studies and example below. These are summarised here.

Table 3: Additional risk-presenting uses of AI in the Southeast Asian region

Use of AI (SE Asia)	Example use	Possible risks
Policing of sex trafficking	Use of AI to monitor signs of trafficking on social media	Blurring between prostitution and trafficking, lack of protection of women's privacy
Social media content recommendation systems	General use of social media for inter-personal and business communications	Propagation of online gender-based violence, misinformation, and polarisation
Driver verification	Use of biometric identification to verify drivers for ride hailing	Misidentification errors affecting either the client or the driver
Improvement of agricultural yield	Use of AI and sensor devices to improve yield in rice farming	Exclusion of female land tenants, leading to greater inequality between farmers

Based on the mapping of national high-risk uses of AI, the committee or task force would then develop a gender or societal risk assessment methodology which would allow for a more systematic and quantitative understanding of risks before deployment. Mitigation measures could therefore be included into the design of the AI systems before deployment, or, in certain cases, other non-AI solutions could be recommended.

The diagram below shows the workflow of a gender-risk assessment process for a high-risk AI system.

Figure 7: Gender or societal impact assessment in AI systems
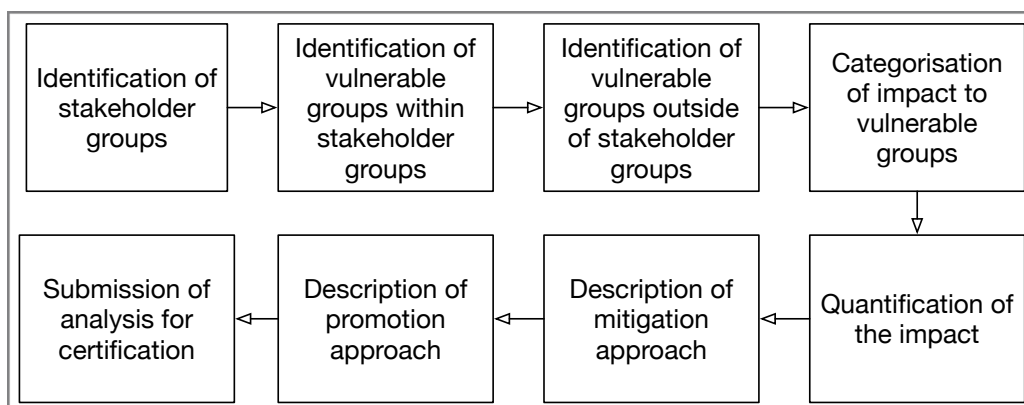
OPPORTUNITY # 2: CREATION OF GENDER AND AI GUIDELINE

The second policy opportunity for the gender and societal impacts of AI is the creation of a normative guideline or module which aims to support private and public sector projects to development gender-safe products.

Depending on the context of the country, there are two possibilities here:

1. Ensuring the inclusion of in-depth gender and other societal analyses in new AI ethics guidelines

2. Developing a gender module for a pre-existing AI ethics guideline, along with additional modules for other disadvantaged groups (such as indigenous communities, youth, the elderly)

Whether or not a distinct AI ethics framework already exists, it would be relevant to include specific sections for groups that could suffer from discrimination, stereotyping or exclusion.

This description would guide both the public and private sectors in considering gender-sensitivity, along with other societal impacts, in their AI system development. It would have the added effect of encouraging AI development teams to hire gender experts, for example, and conducting internal audits before deployment, even if their deployment is not under consideration by the AI ethics commission.

An example of such a guidance is UNICEF's AI for Children Guidance[117], which goes into depth explaining the risks and opportunities of the technology for this specific group. UNESCO[118] has also published a report on the findings of its Global Dialogue on AI and gender equality, which contains key recommendations on integrating gender considerations into AI guidelines. These are summarised below (p.16, Content of Principles):

---

[117] UNICEF (2021). Guidance on AI for Children. Available at: https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf

[118] UNESCO (2020). Artificial intelligence and gender equality: key findings of UNESCO's Global Dialogue. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000374174

Table 4: Links between Gender and AI guidelines and existing global norms

Recommendation (UNESCO, 2020)	Link to opportunity #2
Where Gender is Located: gender equality as a stand-alone principle	A link should be made with existing gender equality initiatives, with support from government agencies focused on women and families, for example.
Whole of Society, Systems, and Lifecycle Approach	Identifying the groups most likely to be discriminated against in AI requires consideration of which groups are already discriminated against or marginalised in each country.
Addressing Gender Equality in AI: Avoiding Harm, Increasing Visibility, and Contributing to Empowerment	This involves explaining risk to women and how to mitigate it, ensuring that datasets are gender disaggregated for better testing and monitoring of bias, and consider uses of AI for women's empowerment.
Provision for Differentiated Impact on Women, Girls and Points of Intersectionality (and multiple forms of discrimination).	Mapping how different uses of AI should be analysed to understand the impact on women and other societal groups, in terms of discrimination, stereotyping, and exclusion.
Provision for participation of Women/Girls and Gender Equality Experts	Engaging national gender equality experts in the development of policy guidance.
Prioritization and Tradeoffs	Considering situations in which the use of AI may be too risky, and alternative solutions might be sought.

OPPORTUNITY # 3: AI SAFETY STANDARDISATION ACROSS THE REGION

This opportunity involves the development of technical standards involving all steps in the AI lifecycle, such as gender-disaggregated datasets, testing and monitoring of AI systems

AI can exclude women and marginalised groups at many different points in its lifecycle, from the use of training datasets to the accessibility of tools and outputs. The table below presents these different entry points in the AI lifecycle. From a standards perspective, there are technical opportunities for risk mitigation at each one of these stages.

Table 5: AI safety standardisation for gender and societal impacts

Stage of the AI development lifecycle	Possible risk categories	Example risk mitigation techniques
Project development and procurement	Exclusion (overall lack of consideration for gender impact in project, and lack of female participation in project)	Gender-impact assessments, gender balance metrics in AI projects.
Data preparation	Discrimination (unequal representation of women in data points, leading to bias)	Use of gender-disaggregated datasets, using an equal amount of data points representing men or women, considering representation of other societal groups in data.
Parameter selection	Stereotyping (allowing the model to select parameters, or indicators, that reflect historical gender biases)	Manual exclusion of parameters known to stereotyping women, testing of models for gender stereotypes
Model training and testing	Discrimination (lack of monitoring for unequal effects on women)	Ensuring adequate testing of model on data points representing women; for example, developing a standard test set which would represent the gender and demographic diversity in Southeast Asia.
Deployment and use	Exclusion (unequal share in benefits of AI system)	Assessing access for the AI system by women
Post-deployment monitoring	Discrimination (lack of monitoring for unequal effects on women, even if these were not apparent during testing)	Ensuring adequate and ongoing testing of model on female users or beneficiaries; ensuring and availability of a safe channel to report discrimination incidents.

At the moment, there are opportunities for national standards bodies to have considerable impact at a regional and international level, as work on gender and societal impacts of AI are still ongoing. For example, the IEEE Algorithmic Bias Working Group includes gender as one of the possible groups that could suffer from discrimination in AI systems[119]. Similarly, the ITU Digital Regulation Handbook on AI[120] includes several

---

[119] IEEE (n.d.) Algorithmic Bias Considerations. Available at: https://standards.ieee.org/ieee/7003/6980/

[120] ITU (2021). Digital Regulation Handbook on AI. Available at: https://digitalregulation.org/3004297-2/

references to gender in its section on algorithmic bias. There are also two relevant standards from the International Organisation for Standardisation (ISO), namely ISO/IEC TR 24368:2022, which refers to Information technology — Artificial intelligence — Overview of ethical and societal concerns; and ISO/IEC DIS 25059, which refers to Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems.

At a regional level, ASEAN has also included AI standardisation as an objective of it's 2025 Digital Master Plan, notably in EA2.7, which reads: "EA2.7: Adopt regional policy to deliver best practice guidance on AI governance and ethics, IOT spectrum and technology".

In addition, national initiatives in AI ethics and standards can be shared across the region. For example, Singapore Government has developed AI Verify, a technical framework which "*verifies the performance of an AI system against the developer's claims and with respect to internationally accepted AI ethics principles.*"[121] Just like other policies discussed in this report, it serves as a general reference discussing bias and discrimination on a general level, however can serve as a starting point for more specific testing of gender and other types of biases.

Current efforts in AI standardisation provide room for countries to develop their own standards which incorporate technical solutions to societal risks of AI. These national standards can also be shared at a regional level in the exchange of best practices and common approaches to AI risks in the private and public sectors.

OPPORTUNITY # 4: INVESTMENTS IN WOMEN IN AI NETWORKS, TRAINING, COMPANY DEVELOPMENT AND POLICY PARTICIPATION

Finally, the last opportunity of this report reaffirms ongoing efforts to accelerate the participation of women in various stages of AI development, such as software development, company ownership, financing and policy development. This includes a number of initiatives, many of which already have some progress in the Southeast Asian region.

---

[121] OECD Policy Observatory (2022). Singapore's A.I. Verify builds trust through transparency.
Retrieved at: https://oecd.ai/en/wonk/singapore-ai-verify

Table 5: Activities supporting women's representation in AI

Activity	Description
Investment in Women in AI and STEM networks	Development of networks allowing women to meet each other and share mentorship and best practices in the STEM and AI fields.
Pipeline development	Support of women's company leadership by charting leadership programmes and other paths leading to top-level positions.
Training and upskilling	Investing in STEM university education, increase the number of AI professors in universities that are women, as well as post-graduate training for those already in the workforce.
Investment in women-led AI companies	Development of venture capital or other types of funding aimed at supporting women-led companies.
Support for women's participation in AI policymaking	Promotion of women's participation in ongoing AI policymaking.

CONCLUSION

In this report, we presented an overview of gender risks in AI and how they might be addressed in ongoing AI policy. The objective of this work was to provide pathways for the better inclusion of women and marginalised groups in Thailand, Indonesia, Malaysia and the Philippines in AI development in these countries.

Overall, there have been very promising developments in AI policies in each country, and throughout the region. Many of the use cases and strategies discussed here are relevant at a regional Southeast Asian level, as ASEAN also works on developing norms and strategies for the region. This work therefore also relates to the ASEAN 2025 Digital Master Plan, which aims to foster "*ASEAN as a leading digital community and economic bloc, powered by secure and transformative digital services, technologies, and ecosystems.*[122]"

According to the 6th Association of Southeast Asian Nations (ASEAN) Summit on Women's Empowerment in the Digital Age, held in June 2020, ASEAN leaders reiterated their commitment to the 1995 Beijing Declaration and Platform for Action and the United

---

[122] Association of Southeast Asian Nations (2020). ASEAN Digital Master Plan. Retrieved from: https://asean.org/book/asean-digital-masterplan-2025/

Nations Security Council Resolution 1325 on Women, Peace and Security[123], to promote gender equality and empowerment of all women, towards realising an inclusive, people-oriented, people-centered ASEAN Community. The commitment to gender equality is particularly important as women continue to disproportionately experience poverty, discrimination, social economic and political exclusion and certain types of violence across both developed and emerging countries[124].

Globally, this work also aims to feed into efforts by United Nations member states to implement the Recommendations on the Ethics of Artificial Intelligence spearheaded by UNESCO. As we have seen, specific analysis of gender risks and opportunities in AI are still lagging, and could be integrated to policy development in any member state. Some of the examples detailed in this paper are region-specific, such as the use of AI for smart rice farming; while others, such as the consideration of the protection of women in ride sharing apps, can be universal. While applications of AI norms will be specific and unique to each country, both artificial intelligence and human rights are universal, opening the door for much exchange and dialogue between countries. It is therefore our hope that this report will inform not only policy dialogue in the focus countries, but provide guidance for AI policy implementation globally.

Next steps in this research may involve a consideration of other societal risks, such as discrimination related to linguistic or other types of minorities in the focus countries. Additionally, considerations related to sustainability and the environment could also be explored further.

---

[123] ASEAN (2020), 'Chairman's Press Statement of ASEAN Leaders' Special Session at the 36th ASEAN Summit on Women's Empowerment in the Digital Age', 26 June, Viet Nam

[124] Sey, Araba. "Gender Digital Equality Across ASEAN." (2021).