How Asean Can Feed A.I. the Right Data

Rebecca Isjwara
The Hong Kong University of Science and Technology

Executive Summary

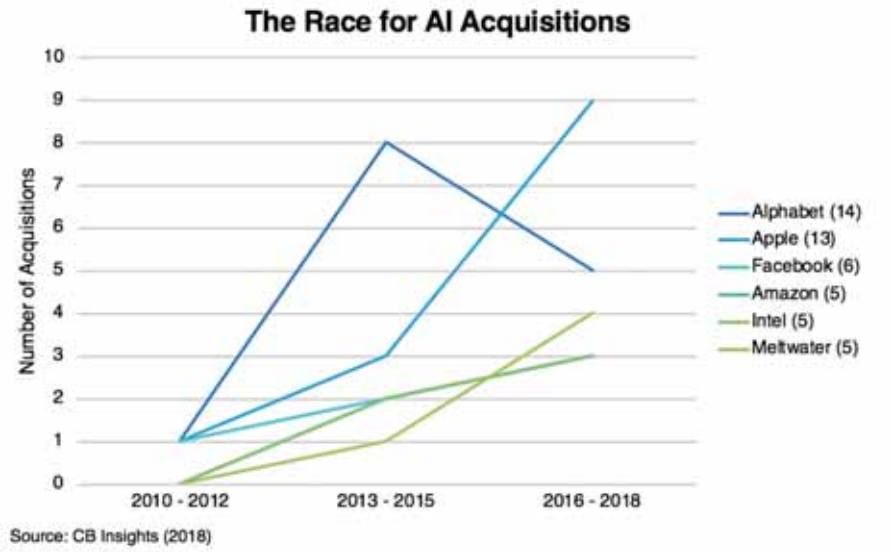
Artificial intelligence (A.I.) development is on the rise, and regulators are scrambling to ensure that the algorithms are fair and responsible. Although the United States and China are two large economies that are making strides in this field, Asean still has some lead time to draft and implement the appropriate regulations and policies before its smartest A.I. developers begin to gain traction.

Currently, there one crucial problem with A.I.: they are only as smart as the data they feed on. The data sets that are presently used by most—if not all—A.I. developers are still biased and do not represent the population fairly. Biased data lead to biased algorithms, eventually yielding biased outputs. In order to prevent Asean from going down this problematic spiral, this problem needs to be addressed at its roots: the data itself. This policy brief recommends a three-prong approach for Asean to feed A.I. more representative data.

A map depicting Asean's member states. Source: United Nations (2012)

The Situation

A.I. is far too prevalent for regions such as Asean to undermine its existence. PWC estimates that A.I. would make up $15.7 trillion of the global economy by 2030[1]. Technology giants have been on an acquisition spree: Alphabet acquired 14 companies since 2010, followed by Apple with 13 and Facebook with 6[2].



Although the development of the A.I. field in the industry is arguably beneficial for world economies, A.I. is only as good as the data it parses. The problem of biased data is crucial as the world is relying more and more on data-based decision making. This fundamental flaw could have serious repercussions for people affected by the outcome of the aforementioned algorithms. UC Berkeley's Deirdre Mulligan encapsulates it simply: "The data isn't fair."[3]

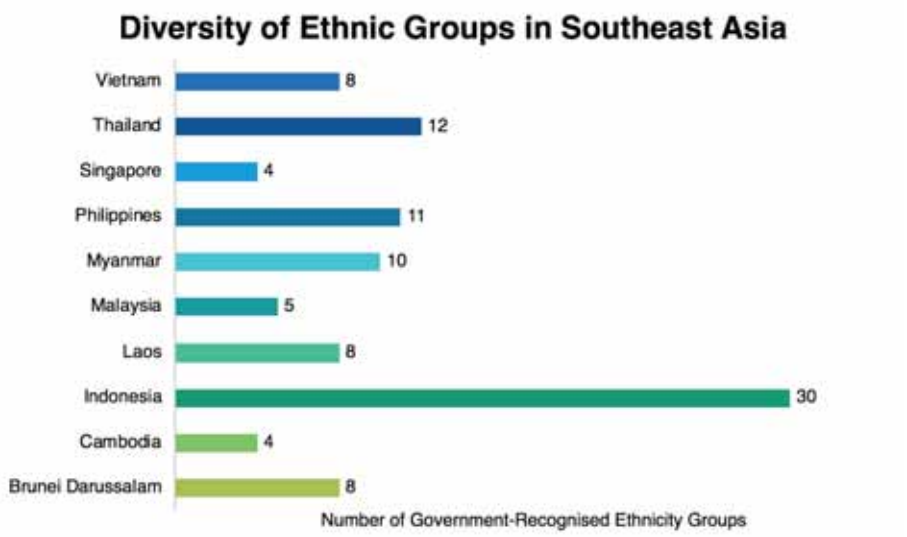
The implications of algorithmic bias can be life-changing for certain marginalised demographics that are not represented fairly in the data used to train A.I. systems. From calculations of credit ratings to the severity of prison sentences to the pricing insurance packages, the numerical outcomes from these algorithms could yield prejudiced results depending on how the algorithm identifies you, the data input. Microsoft's Kate Crawford said that biased data may, in the future, be "influencing our core social institutions."[4]

[1] *PwC (2017). PwC's Global Artificial Intelligence Study: Sizing the Prize. [online] PwC. Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.*

[2] *CB Insights (2018). The Race For AI: Google, Intel, Apple In A Rush To Grab Artificial Intelligence Startups. [online] CB Insights Research. Available at: https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/.*

[3] *Vanian, J. (2018). Unmasking A.I.'s Bias Problem. [online] Fortune. Available at: http://fortune.com/longform/ai-bias-problem/.*

[4] *Ibid.*



Source: United Nations (2001), United Nation Population Fund (2013), Badan Pusat Statistik (2010), Lao Front for National Construction (2006), Department of Information, Ministry of Communications and Multimedia, Malaysia (2015), Population Council (2013), Philippine Statistics Authority (2000), SingStat (2017), Office of the National Culture Commission (2004), Trang Web Thủ tướng (2014)

Recommendations

This policy brief recommends the Asean committee to set up a task force dedicated to mitigating risks stemming from biased data within the region. The task force is delegated a three-prong approach to address the issue in the Asean region.

1 **Create Representative Datasets**: the task force's main responsibility is to create a fully-inclusive data set for algorithms to learn from. This comes from a background of rich racial and ethnic diversity present in the region as well as mobility that allows citizens to travel freely between countries[5]. The datasets curated would be available to the public to encourage input and assure fairness, but more importantly, these datasets could be used by A.I. developers around the region to train their algorithms better. Public participation in curating representative data is also encouraged by inviting participants to submit selfies or take photos at government-run events such as the Asian Games or the biennial Southeast Asian Games.

2) **Auditing Biased Datasets:** the next responsibility of the task force is to offer auditing services to demographic datasets. A common method to overcome biased data is oversampling, which is adding weight to underrepresented elements in a dataset[6]. The task force is also encouraged to be open to auditing suggestions from the public, especially in addressing issues from demographics that are normally underrepresented.

[5] *ASEAN (2006). ASEAN Framework Agreement on Visa Exemption. [online] ASEAN. Available at: http://agreement.asean.org/media/download/20160831072909.pdf.*

[6] *Pew Research Center (2018). Oversamples. [online] Pew Research Center for the People and the Press. Available at: http://www.people-press.org/methodology/sampling/oversamples/9/ [Accessed 17 Jul. 2018].*

3) **Encourage Participation from More Diverse Talent:** one of the most effective methods to discourage biased codes and algorithms is to have a diverse team build them. The third responsibility of the task force is to promote diversity in A.I., technology start-ups, or organizations that work with A.I.. This can be achieved through creating programs or spearheading initiatives that target demographic groups who are not yet as heavily involved in A.I. development.

Conclusion

Feeding A.I. representative data would yield less biased codes and algorithms that would ultimately benefit the Asean community. The Asean region is recommended to set up a task force to take a three-prong approach in mitigating issues related to biased data before they hit the region. The most important responsibility of the task force is to create fully-inclusive datasets that will be available to the public for A.I. development in the region. The task force will also audit data sets to counter the biases present in readily-available datasets. Lastly, the task force should make a push towards diversity in A.I. development teams all across the region. By fully utilizing the lead time the region has, Asean has the chance to ensure a more responsible and fair A.I. industry that would benefit the region as a whole.