The Swedish foundation for human rights
Student Essay Competition 2022
Isak Thorin
Word count: 3997

# See no evil, hear no evil, speak no evil

---

Analysing the human rights legal framework for disinformation legislation and content moderation in ASEAN

# Table of contents

1 **Introduction** ........ 1
1.1 Background ........ 1
1.2 Purpose ........ 2
1.3 Delimitations ........ 2
1.4 Outline ........ 2

2 **Disinformation** ........ 3

3 **Laws against disinformation** ........ 4

4 **Content moderation** ........ 6

5 **Concluding remarks** ........ 8

6 **List of references** ........ 12

# 1 Introduction

## 1.1 Background

Why is southeast Asia important to the study of disinformation? In 2018, Katie Harbath, a Facebook executive, called the Philippines "Patient Zero" in the so-called "global misinformation epidemic".[1] She was referring to how techniques of media manipulation were first observed in the 2016 Philippine presidential election, mere months before Brexit and Donald Trump's election. In addition, in its final report in 2019, the independent fact-finding mission on Myanmar (UN), found that Facebook and social media had played an extensive role in distributing speech that fuelled social attitudes, intolerance and violence against Rohingya,[2] contributing to the conditions under which crimes against humanity were carried out against the Rohingya ethnic minority. [3] Also, in 2022, Ferdinand Marcos Jr. was elected president of the Philippines following an election campaign that observers called a "tide of falsehoods and social-media lies".[4] Perhaps most notably, his campaign aimed to rewrite the very history of the Philippines to influence voters.

This persistent use of social media and the internet in southeast Asia for the spread of disinformation is not by accident. According to a survey from 2021, southeast Asia accounts for 9,9 % of the world's total internet users.[5] These 9,9 % are also some of the most active internet users in the world. According to the same survey Malaysia, Indonesia, Thailand and Singapore all spend more than 8 hours on line, compared to a global average of 06:54 hours per day.

Both actual events, and internet use in the region, therefore show why southeast Asia is important to the study of disinformation. In southeast Asia, disinformation has begun to have serious, real-world consequences for the security of people and possibly states.

---

[1] Rappler, *360/OS Facebook's Katie Harbath on protecting election integrity*, https://www.youtube.com/watch?v=dJ1wcpsOtS4&t=76s [accessed 2022-10-02].

[2] *Detailed findings of the independent international fact-finding mission on Myanmar*, A/HRC/42/CRP.5, para 466.

[3] A/HRC/42/CRP.5, para 211-213.

[4] Papa, M, *Philippines election: "politicians hire me to spread fake stories"*, https://www.bbc.com/news/blogs-trending-61339293 [accessed 2022-10-02].

[5] Kemp, S., *Digital 2021: global overview report*, https://datareportal.com/reports/digital-2021-global-overview-report [accessed 2022-10-02].

## 1.2 Purpose

The rise of the use of disinformation also gives rise to the use of countermeasures against disinformation, both from states and social media companies. Countermeasures against disinformation can take many shapes, one of them legislation or norms. Since disinformation is another form of information, limiting disinformation is to limit a person's right to share and access information. Such rules therefore clearly affect the freedom of expression. There is therefore a clear connection between countermeasures by governments and social media companies that limit or control the spread of disinformation, and human rights. The purpose of this essay is to analyse the human rights legal framework for disinformation legislation and content moderation within the ASEAN region.

## 1.3 Delimitations

This essay will only cover legislation enacted by countries and content moderation by social media companies. This is because these are examples of regulation directly targeting a person's right to spread and receive information. Other initiatives against disinformation such as fact checking or the spreading of "social media awareness" do not directly target the rights of individuals.

Since this essay aims to review the human rights legal framework for countermeasures, this essay will review systems for human rights and not the individual circumstances of an individual country or company. The system I have chosen to review is the ASEAN human rights declaration (AHRD). This essay will only look at the freedom of expression.

## 1.4 Outline

This essay consists of three parts. Part 1 is an introduction where the background of the issue, the purpose, delimitations and general outline of the essay is explained. Part 2 consists a study of the human rights' legal framework in the ASEAN region by studying the impact of two countermeasures against disinformation used in the region. Part 3 is a summary of all conclusions and a discussion about countermeasures against disinformation and the human rights legal framework in the ASEAN region.

# 2 Disinformation

Two fundamental questions regarding disinformation are (1) what it is and (2) why it's dangerous. Disinformation is false or manipulated information, created or disseminated deliberately, intended to deceive or mislead audiences, either to cause harm or to further personal, political or financial gain.[6] This distinguishes it from *misinformation*, which is information that is false, but not created with the intention to cause harm.[7] "Fake news" can be defined as "information that mimics news media content in form but not in organizational process or intent".[8] "Fake news" can therefore become disinformation in e.g., an election when it's used to further a candidate or a political goal.

The exact danger from disinformation is debated. It might be more accurate to say that disinformation can be dangerous in particular circumstances when access to information is critical. A report from *The world economic forum* suggests that one such circumstance might be situations of high tension.[9] The idea is that disinformation could led people to make decisions in high-tension situations that lead to irreparable damage. One example of this could be the events in Myanmar, where disinformation was used to motive actions against the Rohingya in a situation of great tension.

Another circumstance where disinformation could be dangerous is elections. Even though there is no consensus on how well-informed citizens need to be for democracy to work, there is an agreement that citizens need to be at least reasonably informed about politics and society.[10] The worry is that disinformation will lead to the election of representatives on "false premises", which could be detrimental to the voter and society as a whole. The idea is that voting is so important for democracies that intervention into the election process makes a society unstable. An example of disinformation in elections could be the 2022 presidential election in the Philippines. The effects of this election have yet to materialise.

[6] Human rights council, *Role of states in countering the negative impact of disinformation on the enjoyment and realization of human rights*, A/HRC/RES/49/21.

[7] UNESCO, *Journalism, fake news & disinformation: handbook for journalism education and training*, p. 7.

[8] Bader, M., *Disinformation in Elections*, p. 26.

[9] Howell, L., *Digital wildfires in an hyperconnected world*, https://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/ [accessed 2022-10-02].

[10] Strömbäck, J, Wikfors, Å., *Knowledge resistance in High-choice information environments*, p. 3-4.

# 3 Laws against disinformation

Myanmar, Malaysia and Singapore are all states within ASEAN that have legislation aimed at disinformation or are drafting such legislation. Each of these countries have also received criticism on the human rights implications of the legislation.[11] The criticism stems from a worry that these laws are abuses of power. Mainly two things have been highlighted as problematic from the perspective of human rights: the criminalisation itself and the ban on anonymity.

The worry for criminalisation of disinformation becoming a human rights issue stems from a very simple question: what is true? Ultimately this will be a decision by the prosecutor and the courts. If a government controls the judiciary, this opens the way for governments to silence dissidents and political opponents. Basically: if there is an official version of events, these laws may be used to silence any discrepant narratives.

A problem for the enforcement of disinformation legislation is that the online environment is largely anonymous. Prosecution demands that the perpetrator of an alleged illegal act is known. Disinformation legislation therefore often contains a ban on anonymity, or means of recording users and names. However, just as anonymity might be seen as a way to avoid accountability,[12] it can also be seen as a way to protect privacy.[13] The worry from a human rights perspective is that banning or keeping records to "pierce through" anonymity, might be used by an actor seeking to silence opposition.

Given these human rights concerns, the legitimate question of how these laws relate to the human rights system in ASEAN arises. In 2018, the ASEAN countries united behind a framework and joint statement to minimise the harmful effects of fake news.[14] In brevity, this document did not establish any guidelines for the enactment of countermeasures on the regional level, but rather left countermeasures against disinformation a national responsibility. The accompanying declaration to the 2018 framework explicitly recognised that the

[11] Myanmar: Human rights watch, *Myanmar: scrap draconian cybersecurity bill*, https://www.hrw.org/news/2022/02/15/myanmar-scrap-draconian-cy DEWQ N bersecurity-bill [accessed 2022-10-02], Malaysia: Human rights watch, *Malaysia: revoke "fake news" ordinance*, https://www.hrw.org/news/2021/03/13/malaysia-revoke-fake-news-ordinance [accessed 2022-10-02], Singapore: Human rights watch, *Singapore: "fake news" law curtails speech*, https://www.hrw.org/news/2021/01/13/singapore-fake-news-law-curtails-speech [accessed 2022-10-02].

[12] Moore, A., *Anonymity, Pseudonymity, and deliberation: why not everything should be connected*, p. 178-180.

[13] Moore, A., p. 180—181.

[14] AMRI, *Framework and joint declaration to minimise the harmful effects of fake news*, 10 May 2018.

countermeasures must be taken in respect to the state sovereignty and the central role of governments, and the individual circumstances of the different member states. The 2018 framework also encouraged companies and organisations to develop their own industry norms and guidelines against fake news. At the same time, ASEAN also adopted the *Core values of internet literacy for ASEAN*. This document focused on "internet literacy ", meaning how to teach citizens of the ASEAN member states different strategies to become less vulnerable (including to disinformation) on the internet.

These documents form the basis for the ASEAN response to disinformation. These documents show that the ASEAN response to disinformation is based entirely on individuality. It is up to the states and companies themselves to both enact countermeasures and to write guidelines for such countermeasures. The documents do not, neither separate nor taken together, provide any human rights guidelines for such legislation. However, that does not mean that there are no human rights documents in ASEAN with relevance to disinformation legislation. Disinformation laws affect what information is shared or allowed on line, either by criminalisation or restrictions on anonymity. It is therefore clear that article 23 (freedom of expression) in the ASEAN Human Rights Declaration (AHRD) is relevant. The exact scope of article 23 AHRD remains unclear for two reasons, however. The first reason is that there are no secondary sources to interpret article 23 AHRD. The second reason is that the grounds for both restricting and exercising rights in AHRD are very ambiguous. These grounds are found in the general principles of AHRD (article 1-9). These principles become ambiguous through the broadness of their scope. Grounds for *restriction* include: public order, public health, public safety and public morality (article 8). The rules for the *exercise* of rights becomes ambiguous since that are relative to very unprecise goals. Or put otherwise; the enjoyment of rights is always balanced against very broad interests. This is seen in article 6 that states that the rights of the individual must be exercised with regard to the duty of that individual to other individuals and the community and society where one lives. Article 7 AHRD further states that the *realisation* of human rights must be considered in the regional and national context. In conclusion, both the restriction, exercise and realisation of rights within AHRD are unclear.

Lastly, AHRD states that it is the primary responsibility of the ASEAN member states to promote and protect all human rights and fundamental freedoms (se article 6).

In conclusion, this shows that the ASEAN human rights legal framework offers nations complete autonomy to enact disinformation legislation. Not only are they, according to the framework, the actors in charge of enacting disinformation countermeasures, they are also, according to AHRD, the ones to rule the scope of freedom of speech for that same legislation. Human rights are not only balanced against the multitude of interests (susceptible to a wide array of interpretations) expressed in the general principles of AHRD, the very act of making that balancing is done by the individual states.

# 4 Content moderation

Social media companies also write rules of what content is allowed on-line. In the ASEAN region specifically, the 2018 ASEAN framework makes it very clear that social media companies develop these rules on their own.

The term "Content moderation" is used to signify the practice by which social media companies decide (moderate) what content is available for the users on their platforms to interact with.[15] Exactly what information is removed is decided by "platform law". This "platform law" is made up of the website's own terms of service and community guidelines.[16] In other words, "platform law" are rules written *by the companies themselves* for the use of their product (social media or websites). Firstly, this means that there are no specific rules for any region of the world, since content moderation is approached with the starting point of the platform (product), and not of the jurisdiction in which the users find themselves. Secondly, this means that these rules are therefore not strictly speaking laws in the eyes of human rights, since the social media companies are not governments. This is a problem from a human rights perspective since the difference in practical terms between a social media company and a government removing content is non-existent from the individual's point of view. The end result is the same: the information is suppressed and an individual's right to express themselves have been denied.

---

[15] Gillespie, Tarleton., *Platforms, content moderation, and the hidden decisions that shape social media*, p. 5.

[16] *Report of the special rapporteur on the promotion and protection of the right of freedom of opinion and expression*, A/HRC/38/35, para 1.

What is perhaps most troubling from a human rights perspective is that the "platform law" that guides content moderation is not clear. Content moderation is usually performed on an ad hoc basis[17] with vague rules.[18] In other words, content moderation is carried out without an overarching framework. Given the huge impact of content moderation on the freedom of expression, this is less than ideal.

I would argue however, that there are grounds for adopting a human rights approach to the creation of "platform law". The idea is not to start treating companies as governments in courts or commissions, but rather adopt a human rights approach to the creation of "platform law" so that the rules created meet human rights standards.

A starting point for a human rights-based approach to platform moderation could be the *United nations Guiding principles on business and human rights* ("UNGP"). UNGP outlines a non-binding corporate responsibility to respect human rights and to address adverse human rights impacts with which they are involved (article 11). UNGP states that the responsibility to respect human rights is a global standard of expected conduct for all business, wherever they operate.[19]

Since social media companies are internationally active. ICCPR could therefore be a starting point for such an approach. According to article 19 and 19(3) ICCPR, this means that content moderation guidelines would have to respect the demands of legality, legitimacy and necessity. Generally speaking, this would mean that "platform law" must be accessible and clear enough for users to adapt their behaviour accordingly[20] and that restrictions can only be made for certain purposes and that they must a have a genuine link to a factual problem.[21]

Social media companies are also active nationally, however. This means that their users are within the jurisdiction of different human rights systems. Does this mean that there has to be a specific set of rules for each human rights system? Another issue is the fact that the impact of disinformation will also be different in different regions depending on e.g. local historical, societal and religions factors. In other words: the outcomes of "platform law" cannot be the same all over the world. I would argue that this does not necessarily call for a specific set of

[17] Sander, B., *Freedom of expression in the age of online platforms: the promise and pitfalls of a human rights-based approach to content moderation*, p. 966.
[18] A/HRC/38/35, para 26.
[19] UNGP, commentary under Article 11.
[20] UN Human rights committee, *General comment no. 34*, CCPR/C/GC/34, para. 25.
[21] CCPR/C/GC/34, para. 22.

rules for every human rights system or nation, however. The important thing is that the *outcomes* suit the local circumstances. I suggest that the human rights demand for "necessity" in restrictions of the freedom of speech can be used to guide companies in the application of their rules in different local contexts. Grounds for this can be found in the case *Kim Jong-Cheol v. Republic of South Korea*.[22] To reach its decision that no infringement of the necessity requirement in article 19(3) ICCPR had occurred when removing an opinion pull shortly before an election, the UN human rights committee looked at the "recent historical specificities of the democratic political process" in the republic of Korea, [23] and how vulnerable the election culture and climate had been to political manipulation and irregularities in the past.[24]

This highlights the value of a human right approach to content moderation. Rather than rigid rules, they can be used as a framework to inform an approach to legislation and application. It is not a matter of perfectly defining a right, but making sure that the decisions regarding that right follows a procedure fulfilling human rights standards.

# 5 Concluding remarks

Disinformation can be linked to events threatening peace and security in the ASEAN region. Perhaps most notable is the example of the crimes against humanity committed against the Rohingya in Myanmar. But disinformation can also have an effect on the outcomes of elections, as seen in the Philippines. It is therefore perhaps not strange that states and companies have begun to adapt countermeasures against disinformation. In practice, such countermeasures will restrict the ability for individuals to spread or partake in information and are therefore restrictions on the freedom of expression. The state interest of countering disinformation and the individual's interest of protecting human rights are therefore both activated. This means that it is important to review the legal framework for the protection of human rights, to ensure that we have the tools and framework needed to make informed and balanced decisions.

I have shown that AHRD gives the right to freedom of speech for the citizens of the members states. I have also shown that ASEAN has an entirely individualistic approach to both the

[22] CCPR/C/84/D/968/2001.
[23] CCPR/C/84/D/968/2001, para 8.
[24] CCPR/C/84/D/968/2001, para 8.3.

enforcement of human rights and to the countermeasures against disinformation. This is problematic from a human rights perspective. What is disinformation? What is true and not true? The answer to what is true is often a political one if it's allowed to be. Indeed, political parties often clash on what is true or not. To decide what is true by silencing the debate could be temping for many states and a mayor risk for disinformation legislation. Here, human rights have the potential to make the debate about what is true into a factual one. The test of legitimacy means that a state must prove that the proposed legislation fulfils a legitimate goal and that it has a genuine link to that goal (se e.g. article 8 AHRD). This could be described as a lie-detector test. The problem within AHRD is that it is for the states themselves to make such a test, since there is no independent entity (such as a court) with the competence to do so. With no mechanism for enforcement, the protection of human rights therefore runs the risk of becoming voluntary for states. There is no easy solution to this problem. For a court or council to receive such competence would require the members states of ASEAN to formulate and ratify a treaty to that effect. Given ASEAN's individualistic approach and emphasis on sovereignty this does not seem likely.

The "platform law" written by companies in the ASEAN region is also problematic. In practice, there is no difference in the effect on the freedom of speech if a country enacts a law or a social media company writes a guideline for the removal of certain content. My suggested solution to this problem has been for companies to adopt a human rights approach to content moderation. I mean that such an approach could provide the framework for both the writing and application of rules. It is not a matter of perfectly defining the freedom of speech in all regions of the world, but a way of guaranteeing that the process companies use when reaching their decisions for content moderation fulfil human rights demands. A problem for such an approach is that it is entirely voluntary for social media companies to adopt. It is important to remember that companies are created to make profit, not to promote human rights. Social media companies make their profit by collecting data on their users to either use for targeted advertising or for selling to third parties.[25] This actively incentivizes them to have as much information, from as any users, in as many countries as possible. Being banned from countries because of human rights disputes would likely not lead to greater profits for such companies.

[25] Sander, B., p. 953.

In conclusion, the lack of an enforcement mechanism for AHRD threatens to make the respect for human rights voluntary for states, and adopting a human rights approach for companies is also voluntary. This option for actors to choose whether or not to apply human rights rules is the biggest weakness for the human rights' legal framework in the disinformation context in ASEAN.

However, I would argue that the situation is more nuanced than that. The strength of AHRD might not be as a legal instrument, but rather as a tool for policy. Rather than AHRD being useless, it is rather a question of AHRD not being used. I would argue that for the states, the bigger problem might be the 2018 framework that did not provide any guidelines on human rights in the disinformation context. Indeed, AHRD stresses that it is for the countries to use the declaration, something they did not do in the 2018 framework. Creating a court to enforce AHRD would give an easy legal solution to the problem. But it is not certain that creating a court would have such a large effect in the disinformation context. Disinformation is not just a legal problem, it is a societal one. Letting a court decide on how to approach disinformation, is effectively countries outsourcing that discussion to a third party. Given that disinformation is a societal problem, one could argue that the solution must come from society (the states). The greater problem for the human rights legal framework within ASEAN is therefore rather that there are no guidelines for how to approach legislation, and not the lack of a mechanism for enforcement. There is no reason why AHRD cannot be used to create such guidelines. The larger problem then, is the lack of cooperation between nations. That however, is a problem outside the scope of this essay.

Similarly, I would argue that the problem of corporations not being bound to follow a human rights approach is an oversimplification of the issue. In truth, human rights offer corporations a way out of a problem that is very difficult for them to manage, which the ad hoc nature of current content moderation shows. Perhaps most notably, human rights could help companies in the difficult assessment of local circumstances. I argue that regional human rights systems, such as the AHRD, could also be used to give corporations invaluable insight in to how to approach the scope of different rights in different regions of the world. "Necessity" could also be used to inform companies on how to approach local history etc. (se above). Instead of an ad hoc process, human rights therefore promise to give actual points of reference and a framework for decision making. In my view, it is therefore likely that companies will adopt a human rights

approach in some form. The problem of the for-profit nature of companies is rather if they will stay true to the framework and the outcomes. Or put otherwise: if they will choose to make certain decisions in some cases because doing so will lead to larger profits. It may be more lucrative to sometimes moderate content of certain politically sensitive events and keep a government content, than to spread that information. This is difficult problem to solve and beyond the scope of this essay.

# 6 List of references

**Treaties:**

*ASEAN human rights declaration* (AHRD)

Article 6

Article 7

Article 8

Article 23

*United Nations Guiding principles on business and human rights* (UNGP)

Article 11

Article 11, commentary

*International covenant on Civil and Political Rights* (ICCPR)

Article 19

Article 19(3)

**Resolutions and declarations:**

Human rights council, *Role of states in countering the negative impact of disinformation on the enjoyment and realization of human rights*, 1 april 2022, A/HRC/RES/49/21

ASEAN ministers responsible for information (AMRI), *Core values on digital literacy for ASEAN*, 10 May 2018

ASEAN ministers responsible for information (AMRI), *Framework and joint declaration to minimise the harmful effects of fake news*, 10 May 2018

UN Human rights committee, *General comment no. 34*, 12 September 2011, CCPR/C/GC/34

**Case law:**

*Kim Jong-Cheol v. Republic of South Korea,* Human rights committee, Communication no. 968/2001, CCPR/C/84/D/968/2001, 27 July 2005

**Literature:**

Bader, M., "Disinformation in Elections", *Security and human rights*, 2018, vo. 29, p. 24-35

Gillespie, T., *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media'*, New Haven, Yale University Press, 2018

Moore, A., ”Anonymity, Pseudonymity, and Deliberation: why not everything should be connected*", Journal of political philosophy*, Vol. 26. No 2. 2018, p 169-192

UNESCO, *Journalism, fake news & disinformation: handbook for journalism education and training*, Cherilyn, Ireton, Posetti, Julie (eds), place the Fontenoy: United Nations Educational, Scientific and Cultural Organisation, 2018

Sander, B., "Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation", *Fordham International Law Journal,* vol. 43, no. 4, 2020, p. 939-1006

Strömbäck, J., Wikfors, Å., *Knowledge resistance in High-choice information environments*, New York; Routledge, 2022

**Reports:**

*Detailed findings of the Independent International Fact-Finding Mission on Myanmar*, 16 September 2019, A/HRC/42/CRP.5

*Report of the special rapporteur on the promotion and protection of the right of freedom of opinion and expression*, 6 April 2018, A/HRC/38/35

Howell, L., 2013, "Digital wildfires in a hyperconnected world", World economic forum, [online], no publication date available, https://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/ [accessed 2022-10-02)

**Other sources:**

Human rights watch, 2021, "Singapore: 'fake news' law curtails speech", Human rights watch, [online], 13 January, https://www.hrw.org/news/2021/01/13/singapore-fake-news-law-curtails-speech [accessed 2022-10-02]

Human rights watch, 2021, "Malaysia: revoke 'fake news' ordinance", Human rights watch, [online], 13 March, https://www.hrw.org/news/2021/03/13/malaysia-revoke-fake-news-ordinance, [accessed 2022-10-02]

Human rights watch, 2022, "Myanmar: Scrap draconian cybersecurity bill", Human rights watch, [online], 15 February, https://www.hrw.org/news/2022/02/15/myanmar-scrap-draconian-cybersecurity-bill [accessed 2022-10-02]

Kemp, S., 2021, "Digital 2021: Global overview report", Datareportal, [online], 27 January, https://datareportal.com/reports/digital-2021-global-overview-report [accessed 2022-10-02]

Papa, M., 2022, "Philippines election: 'Politicians hire me to spread fake stories', BBC news, [online], 8 May, https://www.bbc.com/news/blogs-trending-61339293 [accessed 2022-10-02]

Rappler, 2018, "360/OS: Facebooks's Katie Harbath on protecting election integrity", Youtube, [online], 24 June, https://www.youtube.com/watch?v=dJ1wcpsOtS4&t=76s [accessed 2022-10-02]