ASSER INSTITUTE
Centre for International & European Law

Legal Questions Related to the Use of Autonomous Weapon Systems

Dr Berenice Boutin

June 2021

This paper was commissioned by the Advisory Committee on Public International Law (CAVV) in relation to the 2021 Advisory Report on Autonomous Weapons of the Advisory Council on International Affairs (AIV) and the Advisory Committee on Issues of Public International Law (CAVV)

Contents

1. Key Terminology and State of the Debate

Autonomous weapons systems (AWS) are defined as weapons systems which, 'once activated, can select and engage targets without further intervention by a human operator'. Within the category of AWS, **semi-autonomous weapons systems** are defined as a weapons system that 'is intended to only engage individual targets or specific target groups that have been selected by a human operator,' while **human-supervised weapons systems** are 'designed to allow human operators to override operation of the weapon system'.[1] Types of AWS have thus been defined by reference to their respective level of machine autonomy, and corresponding level of direct human control or oversight.[2]

The notion of **meaningful human control** (MHC) has played an important role in international debates. Initially developed and promoted by a number of NGOs,[3] it rapidly gained traction, with several states adopting the view that AWS should always remain under MHC, and ample scholarship exploring the concept.[4] If the notion of MHC progressively crystallized as a key standard to evaluate the legality and morality of AWS, it has remained relatively vague, and there is limited substantive agreement beyond the term itself. There are also disagreements on whether MHC should be required over the weapon systems as a whole, or certain of its critical functions (i.e. selecting and engaging targets), or over individual attacks.[5]

Essentially, MHC requires a sufficient and adequate degree of human control of the operator using an AWS. This usually necessitates a minimum degree of cognitive awareness, allowing the exercise of human agency and leading to an informed decision-making.[6] It is not clear whether the possibility for an operator to override an attack, or the necessity for him/her to approve selected targets and fire the weapon, would be sufficient to meet the threshold of MHC. Besides,

---

[1] US Department of Defense, 'Autonomy in Weapons Systems' (2012) Directive 3000.09, pp. 13-14.

[2] Linell A Letendre, 'Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers' (2020) 96 International Law Studies 22, pp. 278-282.

[3] Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (2013), www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf; Human Rights Watch and Harvard Law School's International Human Rights Clinic, 'Killer Robots and the Concept of Meaningful Human Control' (2016), https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

[4] See, e.g.: Daniele Amoroso and Guglielmo Tamburrini, 'What Makes Human Control over Weapons "Meaningful"' (2019) ICRAC Report; Vincent Boulanin et al., 'Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control' (2020) SIPRI Report; Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53; Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control' (2018) ICRC Blog; Michael C Horowitz and Paul Scharre, 'Meaningful Human Control in Weapon Systems: A Primer' (2015) CNAS Working Paper; Thilo Marauhn, 'Meaningful Human Control – and the Politics of International Law' in Wolff Heintschel von Heinegg, Robert Frau and Tassilo Singer (eds), *Dehumanization of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018), 207-218.

[5] UNIDIR, The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward (2014), p. 2; Article 36, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf, p. 2.

[6] Article 36, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf; Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53; Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control' (2018) ICRC Blog.

the notion of MHC is not yet fully operationalised, and more work is needed to translate the concept of MHC into concrete requirements and develop a 'common understanding on what the human role should look like in practice'.[7]

Human-machine interaction (also referred to as human-computer interaction (HCI)) is a multidisciplinary field of research that developed in the 1970s-1980s as computers became increasingly part of everyday life. It seeks to observe and analyse how human users interact with computers, so as to elaborate design requirements for reliable and intuitive display and control interfaces.[8]

In recent years, the notion of human-machine interaction emerged in the debates on AWS as a potential way forward to ensure and safeguard human control over AWS. The CCW GGE LAWS Guiding Principles and a number of States have adopted the (normative) view that a certain 'quality and extent of human-machine interaction' is necessary to ensure that AWS can be used in compliance with international law.[9]

For the purpose of clarity, it useful to note that the notion of human-machine interaction is related to, but distinct from, the notions of **human-machine teaming** (the strategic aim of combining human and machine capabilities in collaborative '[h]ybrid human-machine cognitive architectures'),[10] and **human-machine integration** (the futuristic idea of the augmented soldier with for instance brain-implanted electronics).[11]

The progressive shift in the debate from focusing on MHC to human-machine interaction could be explained by an increased recognition that **technological mediation** affects human decision making in many ways, and therefore binary notions of control and delegation do not allow to grasp the complexity and nuances of the relationship between human decision-making and machine autonomy. The theory of technological mediation essentially suggests that technological tools mediate our relationship to the world, and our use of technologies shapes and impacts human processes of decision making.[12]

---

[7] United Kingdom Expert paper: The human role in autonomous warfare (18 November 2020) UN Doc CCW/GGE.1/2020/WP.6.

[8] Batya Friedman and Peter H Kahn, 'Human Agency and Responsible Computing: Implications for Computer System Design' (1992) 17 Journal of Systems and Software 7, p. 10 and references.

[9] Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (25 September 2019) UN Doc CCW/GGE.1/2019/3, Annex IV, Guiding Principle (c); United Kingdom Expert paper: The human role in autonomous warfare (18 November 2020) UN Doc CCW/GGE.1/2020/WP.6.

[10] Paul Scharre, 'Centaur Warfighting: The False Choice of Humans vs. Automation' [2016] Temple International and Comparative Law Journal 154, p. 151. See also: Margarita Konaev et al., 'U.S. Military Investments in Autonomy and AI: Costs, Benefits, and Strategic Effects' (Center for Security and Emerging Technology, October 2020), p. 25.

[11] Jacob Parakilas, Are Augmented Humans the Future of War?, 5 May 2021, The Diplomat, https://thediplomat.com/2021/05/are-augmented-humans-the-future-of-war.

[12] Peter-Paul Verbeek, 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research', in Jan Kyrre Berg O. Friis and Robert P. Crease, *Technoscience and Postphenomenology: The Manhattan Papers* (2016), 189-204.

For instance, in the context of complex and partly autonomous systems such as AWS, research on the effect of technology on human cognitive functions has shown that human operators are subject to **automation bias**, whereby system operators develop over-trust in the reliability of the AWS, and tend to follow system recommendations.[13] In that situation, human control that could formally be meaningful becomes superficial.

The concepts of technological mediation and human-machine interaction can also provide a useful analytical framework to assess whether vesting humans with mere approval or override functions is sufficient to ensure MHC. Indeed, AWS and other AI-enabled technologies operate at a speed and scale that goes beyond human cognitive capabilities. If an AWS pre-selects a target on the basis of millions of data points analysed in a split-second, the role of human control or supervision becomes ambivalent. At the moment a soldier is faced with the decision to follow or not an algorithmic recommendation to engage a given target, he/she will have limited cognitively-actionable information to act upon in a time-critical environment.

Importantly, reflecting on human-machine interaction in the context of AWS allows to broaden the scope of analysis and to move the debate towards pre-deployment phases. Indeed, in order to safeguard the possibility of MHC, AWS should be designed and developed in a way that, for instance, reduces automation bias and supports human cognitive functions. In that sense, decades of studies in the field of human-computer interaction can inform international law and policy debates on human control.

Yet, in the current debates, the focus primarily remains on the **direct control of operators over critical functions** (i.e. selecting and engaging targets). The focus on critical functions can be explained by the idea that it is legitimate and desirable to automate certain repetitive mundane tasks (e.g. data analysis), but that critical life-and-death decision-making should remain within human agents.

However, this prime focus on life-and-death decisions at the operational level obliviates the fact that, when it comes to AWS and other AI-enabled technologies, human decision-making can be critically important even when not directly related to operational decision-making on target selection and engagement. Indeed, AWS are conceived, developed and tested by humans. Design and policy choices made at an early stage will have direct implications on whether an AWS can be and is being deployed in a lawful and ethical manner. As expressed in the CCW GGE LAWS Guiding Principles, '[h]uman-machine interaction […] may take various forms and be implemented at various stages of the life cycle of a weapon'.

The focus on direct human control over critical functions also fails to address the fact that direct human judgment might become obsolete in situations where the speed and complexity of technology make it impossible for a human to exercise MHC. Accordingly, this paper suggests

[13] Paul Scharre, 'Autonomous Weapons and Operational Risk' (2016), CNAS Paper, p. 31; Daniele Amoroso and Guglielmo Tamburrini, 'What Makes Human Control over Weapons "Meaningful"?' (2019), ICRAC Report, p 9.

that, in the context of AWS, elements of critical human decision-making might be relocated at the earlier stages of conception, testing, acquisition, and deployment.

The idea of **compliance by design** builds on approaches found in the field of ethics of technology, namely value-sensitive design, which prescribe to identify and integrate ethical values at the stage of design and development of technology.[14] Applied by analogy to the legal domain, the compliance by design approach suggests that AWS should be designed and developed in a way that reflects and incorporate international obligations. This approach thus seeks to promote and ensure compliance with international law by integrating compliance with legal norms as system requirements. It is a forward-looking approach with the goal of minimising the risk of occurrence of violations of international law once deployed.

Although the CAAW Updated Advice is meant to focus on AWS, it is important to note that AI technologies are, and will increasingly be, deployed in contexts that do not necessarily fit the AWS narratives (e.g. decision-support systems,[15] air defence systems,[16] battlefield management systems).[17] Such **military applications of AI** are also raising crucial legal and ethical questions which should be urgently addressed, but the predominant focus on AWS in policy and scholarly debates tends to obfuscate discussions on other military applications on AI. Nonetheless, a shift of focus from AWS to military applications AI can be observed in some US and EU policy instruments.[18]

2. Operationalising the Concepts

The dense conceptual background of debates on AWS have made it difficult to operationalise the different concepts. The overall aim is to implement the existing legal framework (including IHL and IHRL), to promote compliance with international law, and ensure accountability for violations. Accordingly, this paper suggests not to isolate and focus on one of the key concepts, but rather to analyse how the different concepts relate to each other, and how they come into play at different stages.

For the purpose of this paper, the following main and sub-stages are identified.

---

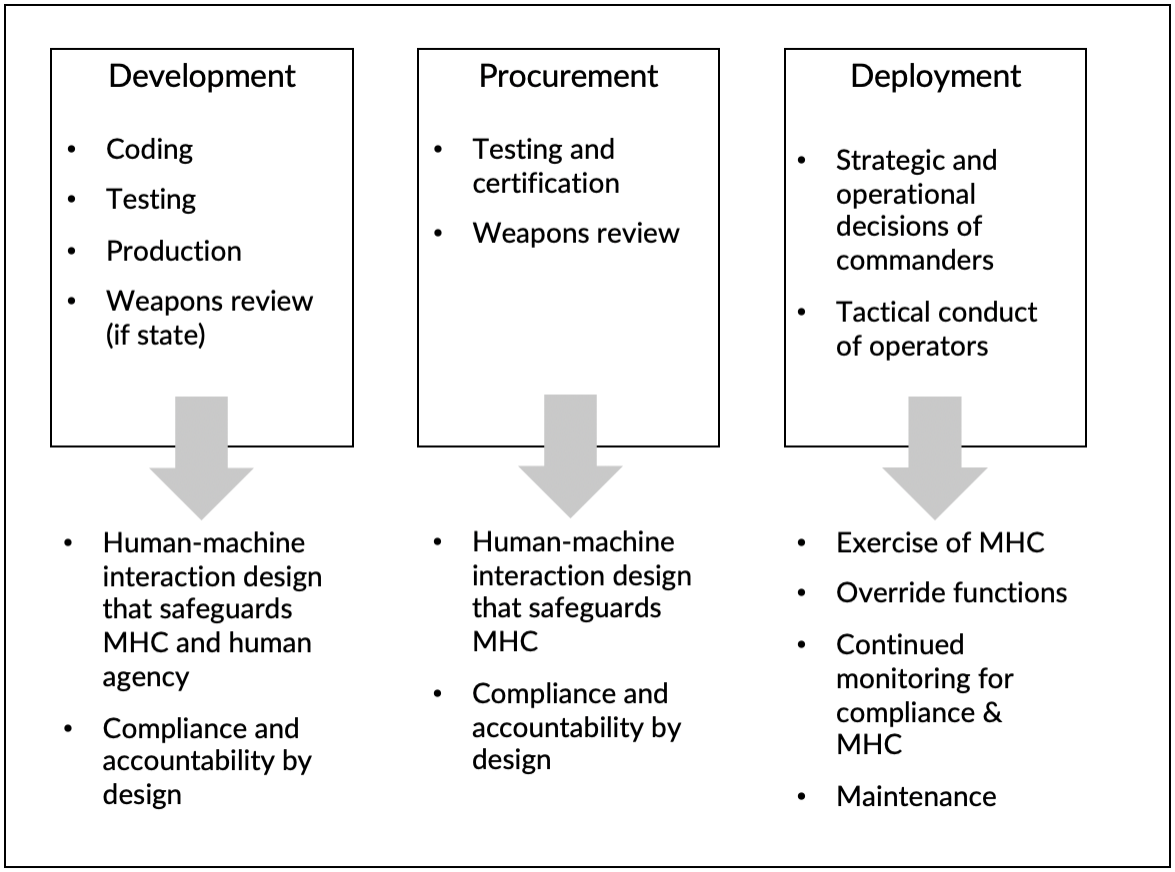
[14] Jeroen van den Hoven, 'Value Sensitive Design and Responsible Innovation' (2013), in R. Owen, J. Bessant, M. Heintz (eds.) Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, 75–83

[15] Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare', (2021) forthcoming in: Yearbook of International Humanitarian Law, available at SSRN.

[16] Ingvild Bode and Tom Watts, 'Worried about the autonomous weapons of the future? Look at what's already gone wrong', Bulletin of the Atomic Scientists, 21 April 2021, https://thebulletin.org/2021/04/worried-about-the-autonomous-weapons-of-the-future-look-at-whats-already-gone-wrong/.

[17] Army Futures Command, 'Project Convergence', https://armyfuturescommand.com/convergence/.

[18] European Parliament resolution of 20 January 2021 on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice (2020/2013(INI)), paras. 3-50 ('International public law and military uses of artificial intelligence'); US Department of Defense, 'Memorandum, 'Implementing Responsible Artificial Intelligence in the Department of Defense', 26 May 2021



At the stage of **development**, design choices will have a significant impact on whether and how an AWS can de deployed lawfully and remain under MHC. It is at this stage that developers can and should seek to ensure effective human-machine interaction, reduce automation bias, secure explainability, and incorporate compliance by design.

At the stage of **procurement**, when a state acquires an AWS from a private corporation or a third state, it has the opportunity to test the system in order to verify and certify that the AWS meets criteria related to design quality that apply at the stage of deployment. Article 36 AP1[19] specifically imposes, at the stage of acquisition of a new weapon, 'an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.' The development and adoption of robust testing and certification mechanisms at the stage of procurement would allow to verify legality under Article 36 and beyond, as well as to assess the quality of human-machine interaction in each system.

At the stage of **deployment**, military commanders take decisions at the strategic or operational levels, which are implemented into action by AWS operators. Commanders are in a position to assess and evaluate a system's reliability and expected performance before taking a decision to deploy, they can request to obtain knowledge on the system technical specifications regarding

[19] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977.

human-machine interaction, and provide mechanisms of continued monitoring of a system performance during deployment. By contrast, operators in charge of firing AWS can have limited control and agency, and cannot be expected to meaningfully control or supervise complex data-driven systems.

3. Forms of Responsibility

The attribution and allocation of responsibility for violations of international law committed with AWS is a fundamental but complex issue. Indeed, there is a multiplicity of actors whose responsibility could be engaged at different stages and under different legal frameworks.

It is important to note that different forms and levels of responsibility are complementary, and that different actors can be found responsible in relation to the same conduct. For instance, the individual criminal liability of soldiers under ICL is concurrent with the responsibility of states for violations of IHL.

The following paragraphs briefly outlines international legal framework of responsibility that are relevant for AWS accountability, and discuss how the characteristics of AWS might complicate allocation of responsibility and potentially lead to responsibility gaps.

State responsibility

At the stage of deployment, state responsibility arises if conduct in violation of applicable obligations is attributable to the state under the ILC Articles on the Responsibility of States for Internationally Wrongful Acts (ARSIWA). States can only act through individuals, and rules of attribution determine which human conduct can qualify as state conduct. In the case of AWS, the main issue is to determine whether, if an operator does not have sufficient control over its conduct, that conduct can nonetheless be attributed to the state. In case of semi-autonomous under human supervision or oversight, attribution of conduct can arguably be grounded in Article 4 ARSIWA.

Besides, state responsibility can arise prior to deployment, for violations of applicable international law obligations committed at the stage of development or procurement. Indeed, a

number of positive obligations of conduct prescribe that states should not only respect but also ensure and promote respect for international law. Additionally, state responsibility is not subject to a requirement of damage, and can arise merely out of a violation. Obligations of diligence applicable at the stage of development include the duty to respect and ensure respect for IHL,[20] and positive obligations to take active steps to secure human rights within a state's jurisdiction.[21] Applied to AWS, it implies a duty to ensure that AI can comply with IHL, to design and train algorithms in line with IHL standards, and to refrain from developing and adopting technology that cannot be IHL-compliant. At the stage of procurement from third parties, there is similarly a duty to verify that AI technology has been designed and developed in line with the obligations of the state. These obligations are at the basis of the compliance by design approach discussed above.

Finally, state responsibility can arise in relation to the conduct of private actors. Indeed, obligations to ensure respect for human rights also come into play in relation to the conduct of private actors within the state's jurisdiction. In order to fulfil this obligation, states must take necessary steps to ensure that private actors respect human rights, including by adopting domestic legislation specifically regulating technological developments.

Individual criminal responsibility[22]

Direct responsibility of operators

For an individual to be principally responsible for the perpetration of a crime under international criminal law (ICL), he/she needs to fulfil the mental elements of *actus reus* and *mens rea* that establish the intentionality of the carried act. *Mens rea* is constituted by (i) intent in relation to the consequences or (ii) awareness of the consequences. However, as explained above, the human soldiers who operate AWS and carry out attacks can have very limited MHC, which would not meet the *mens rea* threshold. As a result, under current ICL, it could be very difficult to hold operators criminally responsible for war crimes committed with AWS.

It has been argued that, under the jurisdiction of the ICTY and in relation to the provision of the Additional Protocol I, the requirement of *mens rea* may be extended to indirect intent, enforcing the risk-taking based mental elements (recklessness and *dolus eventualis*). This approach would allow for the attribution of responsibility to human operators that 'envisage and accept the risk of civilians being attacked'[23] when using or deploying AWS.

[20] Common Article 1 to the Geneva Conventions.

[21] E.g. Article 2 of the International Covenant on Civil and Political Rights (ICCPR); Article 1 of the European Convention on Human Rights (ECHR).

[22] This section was drafted with the assistance of Klaudia Klonowska, Junior Researcher at the Asser Institute.

[23] Marta Bo, Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute (2021) Journal of International Criminal Justice, p. 21.

Commander responsibility

ICL further holds accountable commanders who had (i) a superior-subordinate relationship and effective control over their actions, (ii) knowledge of subordinates' acts; and (iii) who, nevertheless, failed to prevent or punish the commission of the offense. It has been argued that under the command responsibility doctrine, military superiors should be punished for their role in failing to place appropriate limits to the authorized deployment of AWS where those result in unlawful conduct.[24] Under such circumstances, commanders' negligence to actively acquire information and knowledge of the system's performance and consequences of their use may lead to responsibility. It can further argued be that commanders may be held responsible for all conduct of machines that violates IHL if they knowingly decide to activate and deploy an unpredictable AWS, which could be deemed 'reckless under the circumstances'.[25]

Developer responsibility

It is controversial whether developers can and should be held criminally accountable. A programmer who 'intentionally programmes an autonomous weapon to operate in violation of IHL' could certainly be liable for their actions.[26] However, in most cases, the unlawful conduct of AWS will not be a result of intentional errors in programming but the combination of various design choices, human errors, and operational circumstances. The programmer is both physically and temporarily far removed from the operational environment, and engaging their criminal liability beyond intentional wrongful acts would likely amount to scapegoating. In order to promote responsible design, frameworks of domestic civil liability would be more appropriate to the situation of developers.

Corporate responsibility

The responsibility to ensure that corporate activities respect human rights falls primarily on the states, which has a duty adopt and/or enforce domestic legislation to this end. Therefore, corporate responsibility in relation to AWS would primarily be upheld in domestic courts under domestic private law.

Strict liability

Due to the relatively high-risk nature of deploying AWS in conflict, it can be considered whether strict liability would be appropriate. Strict liability is based solely on harm caused, and arise irrespective of negligence or fault. It is well-suited to address technical malfunctions and accidents, but also can provide incentives to make sure that, at the development phase, high

[24] Neha Jain, 'Autonomous Weapons Systems: New Frameworks for Individual Responsibility,' in Nehal Bhuta et al. (eds) *Autonomous Weapons Systems* (Cambridge University Press, 2018), pp. 312-313; Marcus Schulzke, 'Autonomous Weapons and Distributed Responsibility' (2013) 26 Philosophy & Technology 203.

[25] Neil Davison, 'A legal perspective: Autonomous weapon systems under international humanitarian law', UNODA Occasional Papers, No. 30, p. 17.

[26] Ibid.

quality standards are adopted. Strict liability of states under international law would be difficult to achieve as it would require a new treaty; but it has proved a useful approach to inherently high-risk activities.[27] Under domestic law, a strict liability framework applicable to private corporations would need to carefully balance needs for accountability and innovation.

4. Regulatory Options

Rather than seeking to achieve a new treaty that would further regulate AWS or ban certain uses of AWS, regulatory efforts should first focus on implementing and enforcing current legal framework. In particular, states should strive to promote and ensure compliance with their obligations at the pre-deployment stages of development or procurement. In order to facilitate this, interpretative guidance could usefully be developed. Such guiding principles on international law and AWS would elaborate on the content and implications of existing obligations when applied to the novel context of AWS.[28] Such soft instruments are not binding as such but are based on established binding norms and can become authoritative interpretations and/or save as a springboard for new binding norms. Finally, in order to operationalise international law and regulation in this context, non-legal technical standards, verification tools and certification mechanisms, will need to be developed.

[27] See, e.g., in relation to eenvironmental damage resulting from hazardous activity: Draft Articles on Prevention of Transboundary Harm from Hazardous Activities, with Commentaries, 2(2) ILC Yearbook (2001), UN Doc. A/CN.4/SER.A/2001/Add.1, at 146; Draft Principles on the Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities, with Commentaries, 2(2) ILC Yearbook (2006), UN Doc. A/CN.4/SER.A/2006/Add.1, at 106

[28] For instance following a similar process than the United Nations Guiding Principles on Business and Human Rights (UNGP).