Data-driven Learning Systems and the Commission of International Crimes

Concerns for Criminal Responsibility?

Anna Rosalie Greipl*

Abstract

Current discussions on the military use of artificial intelligence (AI), in particular concerning autonomous weapons systems, have largely focused on the challenges for the attribution of individual criminal responsibility for war crimes whenever such systems do not perform as initially intended by human operators. Yet, recent observations evidence the pressing need to shift the discussion on the responsibility gap further to include challenges raised by the intentional use of AI systems for the commission of war crimes and other international crimes. Additionally, the increasing development and use of AI systems, based on data-driven learning (DDL) methods, demands particular attention due to the difficulty these systems' lack of predictability and explainability poses in terms of anticipation of their effects. Against this background, this article complements the present discussion on the responsibility gap by discussing some concerns that the intentional use of DDL systems for the commission of international crimes raises regarding the required mental element and thus, the ascription of individual criminal responsibility. Ultimately, this article proposes preliminary avenues to address these concerns.

1. Introduction

Recent advances in information technologies have brought us to a point where we are increasingly confronted with the existence of artificial intelligence (AI)

* Research Assistant at the Geneva Academy of International Humanitarian Law and Human Rights and PhD candidate at the Graduate Institute, Geneva (Switzerland). I would like to thank Robert Roth, Paola Gaeta, Marta Bo, the other participants in the workshop convened in support of this Special Issue of the Journal, and the *Journal*'s reviewers for their helpful advice and comments. Any errors or omissions are all mine. [anna.greipl@graduateinstitute.ch]

Journal of International Criminal Justice **21** (2023), 1097–1118 https://doi.org/10.1093/jicj/mqad031
 Advance Access publication 4 September 2023

systems, which can act autonomously with little or no human intervention across our daily activities. This development has unleashed a vivid discussion among scholars of law and philosophy on the so-called responsibility gap.[1]

At the international level, the discussions on the responsibility gap continue to focus, essentially, on responsibility concerns in situations where human use or reliance on an AI system results in the *unintended* violations of rules of international humanitarian law (IHL), potentially amounting to war crimes due to errors in, or unpredictable[2] behaviours of, the AI system.[3] In other words, the focus lies on situations where the humans developing or deploying these systems have *no intention* to violate a rule of IHL and commit a war crime. But what if an individual decided to use an AI system with the *will*[4] of committing a war crime or another in international crime? International legal scholars have given little or no attention to this eventuality. Yet, in recent years, a growing and interdisciplinary field of literature, known as 'AI-Crime', has begun to survey and warn about AI systems providing a range of new avenues for criminal exploitation. Most fundamentally, this literature focuses on how humans may use AI systems' capabilities with the *will* to commit or facilitate malicious threats against real-world targets.[5]

Unfortunately, recent observations underline the urgency of these warnings, and attest to the fatal repercussion that the malign use of these systems may produce. For instance, it was reported that the Chinese government uses its rapidly expanding networks of surveillance cameras in combination with AI systems to 'identify Uighurs based on their appearance and keeps records of their comings and goings for search and review'.[6] These are among the first

1 See among others A. Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata', 6 *Ethics and Information Technology* (2004) 175; F. Santoni de Sio and G. Mecacci, 'Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them', 34 *Philosophy & Technology* (2021) 1057; P. Königs, 'Artificial Intelligence and Responsibility Gaps: What Is the Problem?' 24 *Ethics and Information Technology* (2022) 36.

2 The meaning of predictability will be explored in Part 2 of this article.

3 See the article by Dustin Lewis in this Special Issue of the *Journal*. See also Human Rights Watch and International Human Rights Clinic, *Mind the Gap: The Lack of Accountability for Killer Robots*, 9 April 2015, available online at https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots (visited 30 July 2021); M. Bo, 'Autonomous Weapons and the Responsibility Gap in Light of the *Mens Rea* of the War Crime of Attacking Civilians in the ICC Statute', 19 *Journal of International Criminal Justice* (*JICJ*) (2021) 275; T. Chengeta, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law', 45 *Denver Journal of International Law & Policy* (2016) 1; R. Sparrow, 'Killer Robots', 24 *Journal of Applied Philosophy* (2007) 62.

4 The term 'will' is used in this article to refer to the volitional component of the mental element of a crime.

5 K.J. Hayward and M.M. Maas, 'Artificial Intelligence and Crime: A Primer for Criminologists', 17 *Crime, Media, Culture* (2021) 209; M. Caldwell et al., 'AI-Enabled Future Crime', 9 *Crime Science* (2020) 1.

6 P. Mozur, 'One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority', *The New York Times*, 14 April 2019, available online at https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html (visited 14 March 2022).

signs of AI systems deliberately used for the application of a large-scale and automated policy of discrimination that is contributing to what the Office of the United Nations High Commissioner for Human Rights' report depicts as potentially amounting to 'international crimes, in particular crimes against humanity'.[7] In other words, present AI systems can be used intentionally for the commission or facilitation of international crimes.

These observations evidence the pressing need to move the discussion on the responsibility gap further to include the challenges raised by the intentional use of AI systems for the commission or facilitation of international crimes. Moreover, as some scholars have pointed out, the increasing development and use of AI systems based on data-driven learning (DDL) methods demand particular attention due to the seriously unpredictable nature and insufficient explainability of these systems. As we shall see, these characteristics of DDL systems tend to significantly reduce the human capacity to anticipate the outcome of their use. This may, in turn, impact the proof of the mental element when used for the commission or facilitation of an (international) crime.

Consequently, this article intends to complement the present discussion on the responsibility gap by providing an initial examination of the impacts that the malign use of DDL systems has on the ascription of individual criminal responsibility for international crimes. More precisely, the purpose of this article is twofold: first, it examines whether and to what extent the malign use of DDL systems may create gaps in the finding of the mental element, and secondly, should these gaps materialize, it proposes preliminary avenues to address these legal issues.

Yet, considering that the concerns related to the malign use of DDL systems are relatively new in the international criminal law literature,[8] not all relevant questions regarding this phenomenon can be addressed in this article. Therefore, the scope of this article needs to be limited in various ways. First, the present discussion focuses exclusively on the responsibility challenges related to the finding of the mental element (*mens rea*). However, it does not deny the significance that the malign use of DDL systems may additionally have on questions related to the establishment of the objective element (*actus reus*).[9] Secondly, the malign use of AI systems can take innumerable forms, including for the commission or facilitation of international crimes, thus

7 United Nations Office of the High Commissioner for Human Rights, 'OHCHR Assessment of human rights concerns in the Xinjiang Uyghur Autonomous Region, People's Republic of China', 31 August 2022, § 6.

8 One of the few examples where the malign use of AI is addressed is in: G. Fattori, 'Traditional Categories of International Criminal Law Challenged by Technological Progress: Mental Element and Malign Uses of Autonomous Weapon Systems' (LLM thesis at the Academy of International Humanitarian Law and Human Rights, Geneva), on file with the present author.

9 For this article, it will be assumed that the action resulting from the use of or reliance on an intelligent system corresponds to the objective element (*actus reus*) of a crime as foreseen under the ICC Statute. However, it is useful to note that some commentators have questioned the existence of the objective element when AI systems are used, maintaining that the operator does not perform a voluntary act. On this point see for instance: C.W. Westbrook, 'The Google Made Me Do It: The Complexity of Criminal Liability in the Age of Autonomous Vehicles

triggering questions on both direct criminal responsibility for the commission of a crime and other modes of responsibility. This is important to bear in mind because the mental element required for indirect modes of criminal responsibility differs from that of direct commission. While addressing both sets of questions is important, this article focuses on, as a first step to further this debate, situations where the DDL system is used with the intent to *commit* an international crime. In other words, it is concerned with situations where criminal conduct is produced by the use of an intelligent system — for instance, a DDL weapon system used to kill or wound a person *hors de combat*.[10] It does not contemplate situations in which DDL systems are used to *facilitate* the commission of international crimes — for instance, an AI-driven surveillance system used in furtherance of state or organizational policy that consists in directing attacks against a civilian population. Finally, the mental element definition of Article 30 of the ICC Statute serves as a reference point for the present discussion. This should, however, not divert attention from the fact that different definitions of the mental element for the commission of international crimes exist at the international and domestic levels. These definitions may, on the one hand, raise different or additional challenges, and on the other hand, be the source of avenues to reduce the responsibility gap under the ICC Statute's default definition of the mental element.

This article begins by describing the predictability and explainability challenges of DDL systems to highlight the difficulty that they could pose in terms of anticipation of their effects. It then presents two scenarios of malign uses of DDL systems for the commission of an international crime that will serve as reference points in this analysis. The third part explores the ICC Statute definition of the mental element. More specifically, it focuses on the level of human awareness required, regarding the prohibited consequences, for the mental element to exist. The fourth part responds to the twofold objective of this article, drawing on the two scenarios introduced in the second part: it first examines the impact that the malign use of DDL systems for the commission of international crimes has on finding the mental element as defined under Article 30 of the ICC Statute; and secondly, explores some potential avenues to reduce the responsibility gap it has identified. Finally, the conclusion outlines the main findings of this article and reiterates the need to continue the discussion on the responsibility challenges generated by the malign use of DDL systems at the international level.

Symposium: The Transformation of Transportation: Autonomous Vehicles, Google Cars, and Vehicles Talking to Each Other', *Michigan State Law Review* (2017) 97, at 128.

10 Following the definition of Rule 47 of the ICRC Study of Customary International Humanitarian Law, a person *hors de combat* is '(a) anyone who is in the power of an adverse party; (b) anyone who is defenceless because of unconsciousness, shipwreck, wounds or sickness; or (c) anyone who clearly expresses an intention to surrender; provided he or she abstains from any hostile act and does not attempt to escape'.

2. The Rise of DDL Systems and Their Impact on Human Comprehension

Until a few decades ago, most AI systems were essentially deterministic and confined to operating within a very restricted and specific problem space of human-crafted sets of symbols and rules. Although these so-called 'expert systems' continue to be particularly useful in supporting humans working on repetitive problems in well-defined domains, these intelligent systems are characterized by a major limitation: they cannot improve their performance by themselves. To overcome this limitation, new methods focusing on creating systems with the capability to learn by themselves started to be explored.[11]

In recent years, we have seen major breakthroughs in the field of DDL methods, including machine learning[12] and a subfield of it, so-called deep learning.[13] These DDL systems have demonstrated remarkable performance in tasks such as image recognition, speech recognition, and machine translation. Despite all their remarkable achievements, present DDL systems are far from providing solutions to all real-world problems. They come with their benefits as well as drawbacks. Some of their key characteristics — that have been further exacerbated with the introduction of deep learning — come to pose fundamental challenges to the proof of the mental element as provided under Article 30 of the ICC Statute, and thus, the ascription of individual criminal responsibility. This concerns, in particular, two key characteristics of these systems: their reduced level of explainability[14] and predictability. As we shall see, these two characteristics of DDL systems significantly reduce the human ability to anticipate the outcome that the system will produce and to comprehend why the system produced that effect.

Understanding this impact on human comprehension is essential in evaluating the extent to which the malign use of DDL systems resulting in the commission of international crimes challenges the proof of the mental element as provided under Article 30 of the ICC Statute. This is because, as we shall see, to establish the ICC Statute definition of the mental element for crimes that include in their definition a material element of consequence, it may be

11 Note that some parts of this section are taken from the author's preliminary PhD dissertation: A.R. Greipl, 'Artificial Intelligence Demystifying the Human in International Humanitarian Law' (Preliminary Thesis Dissertation at the Graduate Institute, Geneva.)

12 Machine learning refers to a subfield of computer science and AI, which broadly speaking aims at enabling systems to 'learn' autonomously and improve from experience without humans explicitly programming the system's knowledge. Accordingly, the human's role is to collect adequate training data and decide on a machine learning method given a system's particular task objective. See: T.M. Mitchell, *Machine Learning* (1st edn., McGraw-Hill Education, 1997).

13 Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain, so-called artificial neural networks. In addition to more traditional machine learning methods, deep learning allows the system to 'learn' from large amounts of data while selecting the best features within these large data sets on their own. See: I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning* (The MIT Press, 2016).

14 Note that other terms are used in the literature in reference to the concept of explainability, including 'understandability', 'transparency', 'interpretability' and 'intelligibility'.

necessary to prove that the accused had the awareness that the consequence 'will occur in the ordinary course of events'.[15] Hence, the following paragraphs explore the extent to which the two key characteristics of DDL systems impact the human comprehension of the effects generated by their use.

A. Human Comprehension of DDL Systems' Use: Predictability and Explainability Challenges

Predictability is the extent to which a system's effects[16] can be precisely and consistently anticipated.[17] It pertains to the question of *what will the system do*.[18] Most AI systems present a degree of unpredictability even if they exhibit a high level of reliability.[19] This is because, in reality, situations in which the environment is completely known *a priori* to the designer or operator are rare. In other words, it is nearly impossible for humans to anticipate all situations a system will encounter in a particular environment and, therefore, the manner in which the system will respond to these new situations.

DDL methods have enabled the creation of systems with self-learning capability, allowing the systems to compensate for and correct prior knowledge provided by the designer. This capability to learn has proven crucial for systems to navigate a variety of environments. For the development of self-driving cars, for instance, the system's capability to learn is essential to navigating extremely open-ended driving environments in which there 'are no limits to the novel combination of circumstances that can arise'.[20]

At the same time, this ability to learn renders it more difficult for humans to predict DDL systems' produced outcomes. The reason for this is that, after sufficient learning experience in an environment, the behaviour of a DDL system can become effectively independent of its prior programmed knowledge. As a result, a certain level of unpredictability is inherent to DDL systems: 'they demonstrate emergent behaviours that are impossible to predict with precision — even by their own programmers. These behaviours manifest themselves only through the interaction with the world and the agents in the environment.'[21]

15 Art. 30(3) ICCSt.

16 The terms result, effect and outcome are used interchangeably referring to anything that occurs because of some action — and disregarding the results' (un)lawfulness.

17 R.V. Yampolskiy, 'Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent' 7 *Journal of Artificial Intelligence and Consciousness* (2020) 109, at 110.

18 UNIDIR, 'The Black Box, Unlocked' (2020), available online at https://unidir.org/publication/black-box-unlocked (visited 27 March 2021), at 5.

19 Predictability needs to be distinguished from reliability. The latter relates to the extent a system is able to perform its task — under stated conditions and for a specified period of time — without failure.

20 S.J. Russell et al., *Artificial Intelligence: A Modern Approach* (3rd edn., Global edition, Pearson, 2016), at 40.

21 I. Rahwan and M. Cebrian, 'Machine Behavior Needs to Be an Academic Discipline', *Nautilus*, 20 March 2018, available online at https://nautil.us/machine-behavior-needs-to-be-an-academic-discipline-237022/ (visited 2 October 2022).

Moreover, the nature of the environment in which a DDL system is meant to operate will highly influence its level of predictability. Indeed, a system operating in a fully observable, static and discrete environment (such as a chessboard) is likely more predictable than a system that will operate in a partially observable, dynamic and continuous environment (such as most battlefield conditions).

This being said, what is important to understand is that the inherent level of unpredictability may, in some situations, be a weakness of the DDL system while in others, one of its central features. Microsoft's chatbot named 'Tay' released in 2016 as an experiment in 'conversational understanding' offers an illustration of the former. The chatbot was thought to learn to engage with people by doing. But within 24 hours, Tay started tweeting racist and other offending statements. Part of this is because Tay had a built-in mechanism that repeated what Twitter said to her, by recognizing language patterns, but without truly understanding the meanings of the words.[22] On the contrary, DeepMind's AI systems famously beat two champion Go[23] players in 2016 and 2017, 'in part with moves, that experts described as ''alien'' and ''from an alternate dimension'''.[24] Thus, in some instances, the common assumption that unpredictability is essentially wrong may be false. A DDL system may well achieve better results than a human by taking unpredictable actions.

A second characteristic of DDL systems that significantly reduces human certainty about the system's outcome concerns their decreased level of explainability. Explainability is the 'degree to which any given system can be understood by any given person' and relates to the question of *why the system does what it does*.[25] What is important to understand is that explainability is not only determined by the technical features of an AI system but also by the human's capacity for understanding.

Although explainability considerations are important for all types of AI systems, they are particularly challenging regarding DDL systems. Indeed, the currently best-performing DDL systems are often the least explainable ones. The insights about the data and the task these systems solve are hidden in increasingly complex models often described as the 'black box'. For example, with deep learning, a single prediction may involve millions of mathematical operations. It would be impossible for a human to do the exact mapping from the data input to the DDL system's prediction.

Given this difficulty, humans train such systems and observe their behaviour to build a reliable mental model, meaning a solid representation of how the

22 Illustration taken from Greipl, *supra* note 11, at 35. See also: Y. Liu, 'The Accountability of AI — Case Study: Microsoft's Tay Experiment', *Medium*, 16 January 2017, available online at https://chatbotslife.com/the-accountability-of-ai-case-study-microsofts-tay-experiment-ad57701 5181f (visited 27 March 2021).

23 Go is played on 19 by 19 boards which allow for approximately 2.1×10^{170} possible configurations (in comparison, chess has 1050 configurations).

24 Illustration drawn from: UNIDIR, *supra* note 18, at 6.

25 *Ibid.*, at 9.

DDL system works in the real world. This permits humans to simulate a system's behaviour without understanding its decision-making process in detail. While these mental models can be useful for many DDL systems' applications, they will highly depend on the human operators' understanding of the system. This includes aspects such as their technical literacy, 'their knowledge of that system's past performance, their knowledge of the system's training data, their understanding of the environment to which the AI system is being deployed and the data it will ingest, and the level of attention they can give the system in an operation'.[26] This means that a soldier operating a DDL system must, among others, be capable of assessing that the conditions the DDL system encounters on the battlefield do not differ significantly from those during the training phase. Considering that most situations in armed conflict are highly fluid, complex and fast-changing environments, this would arguably be possible only for systems given extremely narrow tasks and operating in very unique battle spaces.[27] Otherwise, it is likely that the operator's mental model is not able to reliably anticipate how the system will respond to the new situations.[28]

In essence, the two key characteristics of present DDL systems presented significantly reduce the human ability to anticipate the system's effects and to comprehend why the system produced those effects.

B. The Use of a DDL System for the Commission of an International Crime: Two Scenarios

This subsection presents two different scenarios of malign uses of a DDL system for the commission of an international crime. These will serve as a basis for the following discussion on the impact that the reduced human comprehension of effects generated by the use of DDL systems has on demonstrating the mental element as defined under the ICC Statute.

A preliminary remark should be made regarding the choice of international crime for the following two scenarios. From a criminal law perspective, the use of DDL systems raises particular concerns in relation to so-called crimes of result — i.e. crimes that prescribe that a certain harmful result must occur as a result of the prohibited conduct. This is because, as we have seen, key characteristics of DDL systems tend to reduce the human ability to anticipate *what the system will do* and to challenge the human's understanding of *why the system does what it does*. In contrast, the use of DDL systems raises fewer concerns for crimes that are fulfilled by the mere commission of a prohibited

26 *Ibid.*, at 10.

27 For instance, an AI-driven weapon system that is tasked with firing at a precise object (such as tanks), for a strictly limited duration, and in an environment where the probability of civilians and persons *hors de combat* being present is non-existent.

28 This predictability challenge of some AI systems, also described as 'inherent operational unpredictability', has by some commentators and organizations been a central reason for objection to the development of lethal autonomous weapon systems.

conduct, also known as crimes of conduct,[29] considering that the mere use of DDL systems is, thus far, not prohibited per se. Therefore, to assess the challenges posed by malign uses of a DDL system, we will take as an example the war crime of killing or wounding a person *hors de combat* in the context of an international armed conflict,[30] which is a crime of result.

The first scenario is the one where, in an international armed conflict, soldier A of State Alpha wants to kill B (a wounded soldier of State Beta lying on the battlefield). A activates a DDL weapon system to kill B and B dies. The scenario illustrates a situation in which an operator uses a DDL system with the will to commit the war crime of killing or wounding a person *hors de combat*, and this use results in the wanted outcome. This scenario has the particularity that nothing observable indicates any deviation from the operator's wanted outcome and the actual outcome. Considering the predictability and explainability challenges of DDL systems, it is unlikely that soldier A did foresee in detail how the DDL weapon system would proceed to kill B. From a mere technical point of view, depending on the operational context and operators' understanding, however, he or she may have had an enough solid representation of how the DDL system works in the real world to anticipate the outcome with a significant degree of certainty.

The second scenario draws on the former but illustrates a situation where there is a deviation from the operator's wanted outcome and the actual outcome: soldier A still wants to kill B (a wounded soldier of State Beta lying on the battlefield). A activates a DDL weapon system to kill B, but C (another wounded soldier of State Beta) dies as a result of the strike. In this case, the crime is the same as the wanted one (killing a person *hors de combat*), but the victim is not the one envisaged by the operator (wounded soldier C instead of soldier B). The underlying assumption here is that the operator was not indifferent to the harm the use of the system would cause. Instead, he or she did aim at a specific victim. Thus, in contrast to the former scenario, there is an observable deviation from the wanted outcome. This visibly indicates that in this scenario, the operator was not able to accurately anticipate the outcome of the DDL system's use.

From a criminal law perspective, these two scenarios of DDL systems intentionally used to commit an international crime raise essential questions: how much does the human — who uses a DDL system with the will to commit an international crime — need to know about the precise consequence and the course of events leading to that consequence to be held responsible? And does

29 Examples of war crime definitions under the ICCSt. merely proscribing a certain conduct include the declaration that no quarter will be given (Art. 8(2)(b)(xii) and (e)(x) ICCSt.), the employment of poison without necessarily injuring a particular person (Art. 8(2)(b)(xvii) ICCSt.) and the conscription or enlisting of children under the age of 15 years into the national armed forces (Art. 8(2)(b) (xxvi) ICCSt.). For further examples, see also Art. 8(2)(b)(xvii–xx) ICCSt.

30 Art. 8(2)(b)(vi) ICCSt. will serve as a reference point for the present analysis.

the legal response to that question vary as to whether the outcome differs or not from the one wanted by the human operator?[31]

3. Mental Element and the Level of Awareness About the Prohibited Consequences

To address these questions, one shall explore the level of awareness a person is required to have about a prohibited consequence and the process leading to this consequence to prove the mental element of the crime. The starting point is the definition of the mental element under Article 30 of the ICC Statute.

The definition demands that the material elements of a crime — that is, conduct, consequences and circumstances — must be fulfilled, 'unless otherwise provided', with 'intent' and 'knowledge'. It thus distinguishes the volitional (element of will) and cognitive components (element of awareness/knowledge) of the mental element that must exist for individual criminal responsibility to arise unless otherwise provided.[32] The conjunctive ('and') rather than a disjunctive ('or') formulation of the definition[33] echoes the dominant view in contemporary international criminal law that, generally, 'one cannot perform an action or cause a consequence intentionally unless one also has knowledge of the circumstances in which that action or consequence was committed'.[34]

Each of the terms 'intent' and 'knowledge' is specifically defined in paragraphs 2 and 3 of Article 30 of the ICC Statute respectively, and with reference to the material elements of the definition of the crime. Article 30(2) of the ICC Statute defines 'intent' with respect to conduct and consequence; while paragraph (3) defines 'knowledge' in relation to circumstances and consequences.

It is widely acknowledged that this definition of the mental element includes at least two gradations, generally known in civil law systems as *dolus directus*

31 As will be discussed further, similar situations of deviations in the prohibited consequence have been addressed in domestic legal systems through the doctrine known as *aberratio ictus* (accidental harm to a person) in civil law or 'transferred intent' in common law systems. The standard example is as follows: A wants to kill B and points his or her weapon at B; but because A shoots badly, he or she does not kill B, but C, who was near B. As in our example, the failure results from the execution. Although the similarities with our second scenario may at first seem evident, the parallelism needs to be made with caution. This is because the use of DDL systems may raise additional challenges in establishing the causal link between the accused's conduct and the proscribed consequence.

32 Decision, *Katanga and Ngudjolo Chui* (ICC-01/04-01/07-717), Pre-Trial Chamber, 30 September 2008, § 529; A. Eser, 'General Principles of International Criminal Law, 23 Mental Elements–Mistake of Fact and Mistake of Law', in A. Cassese, P. Gaeta and J.R.W.D. Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary*, Vol. 1 (Oxford University Press, 2002), at 409.

33 This is the result of a decision taken during the preparatory work of Art. 30 ICCSt., see: K. Ambos and O. Triffterer, *Rome Statute of the International Criminal Court: A Commentary* (3rd edn., Bloomsbury T & T Clark, 2016), at 1117.

34 *Ibid.*

in the first degree and *dolus directus* in the second degree.[35] Moreover, the conjunctive formulation of the definition requires for each gradation the existence of both, a volitional and a cognitive component. What distinguishes these gradations of the mental element is the relative level of these two components.

As we know from the discussion above, present DDL systems tend to significantly reduce the human ability to anticipate the consequences such system produces. This suggests that in situations of malign use of DDL systems, it is first and foremost the finding of the cognitive component that poses a challenge for ascertaining the mental element. For this reason, it is necessary to assess what level of awareness about the proscribed consequences is required for each of the two gradations covered by Article 30 of the ICC Statute.

A. Awareness of the Prohibited Consequence

Article 30(2) of the ICC Statute defines 'intent' with respect to the material element of consequences[36] providing two different levels of gradation. According to the first alternative, a person has 'intent' with respect to a consequence if he or she 'means to' cause the consequence (Article 30(2)(b)). As just noted, it is generally endorsed in case law of the ICC[37] and scholarship[38] that this first level of gradation covers *dolus directus* in the first degree. There are diverging views, however, as to whether the cognitive component of this gradation under Article 30 of the ICC Statute differs from its counterpart in domestic law.

In common law and civil law systems, the law relating to this gradation of intent generally requires only a very low cognitive component accompanied by a very high volitional component. Thus, the accused does not need to be certain that he or she will succeed in bringing about the prohibited consequence. Even the belief that it is very unlikely that he or she will successfully bring about the proscribed consequence would suffice.[39] Most scholars seem to

35 In common law systems, these two gradations of intent are largely known as 'direct intent' and 'oblique intent'. Yet, this article will refer to the two notions generally used in civil law countries, namely *dolus directus* in the first and second degree. See: S. Finnin, 'Mental Elements Under Article 30 of the Rome Statute of the International Criminal Court: A Comparative Analysis', 61 *The International and Comparative Law Quarterly* (2012) 325, at 332.

36 Note that the notion 'the element of consequence' or 'consequence element' of a crime refers to 'either a completed result, such as the causing of death, or the creation of a state of harm or risk of harm, such as endangerment'. Ambos and Triffterer, *supra* note 33, at 1114-1115. The notion of consequence will, thus, be used in its narrow criminal law sense to avoid confusion with terms used more broadly for the technical discussion, such as result, effect, or outcome.

37 See, Decision, *Lubanga* (ICC-01/04-01/06-803-tEN), Pre-Trial Chamber, 29 January 2007, § 351; Judgment, *Katanga* (ICC-01/04-01/07-3464), Trial Chamber, 7 March 2014, § 774; Decision, *Bemba* (ICC-01/05-01/08-424), Pre-Trial Chamber, 15 June 2009, § 358.

38 Eser, *supra* note 32, at 914; K. Ambos, 'General Principles of Criminal Law in the Rome Statute', 10 *Criminal Law Forum* (1999), at 21; Finnin, *supra* note 35, at 341.

39 Finnin, *supra* note 35, at 164.

adopt the view that this same standard of *dolus directus* in the first degree is reflected in Article 30 of the ICC Statute.[40]

Still, some voices to the contrary have maintained that Article 30 of the ICC Statute sets a higher standard when consequence elements are involved because its cumulative wording in paragraph 1 requires the accused to have 'intent' *and* 'knowledge'. Following this view, a crime with a consequence element must be committed not only with 'intent' ('means to' cause a consequence) but also with 'knowledge' (that the person is aware that the consequence would occur in the ordinary course of events). Eventually, this requirement of 'knowledge' elevates the gradation of *dolus directus* in the first degree to a more stringent standard than its counterpart in domestic law. As Finnin points out, 'unlike in domestic law, where this gradation is normally characterised by a very high volitional component and only a low cognitive component, under Article 30 this gradation is characterised by very high volitional and cognitive components (when it is applied to consequences)'.[41]

The second alternative of Article 30(2)(b) of the ICC Statute considers that a person also has 'intent' with respect to a consequence if he or she 'is aware that it will occur in the ordinary course of events'.[42] This second level of gradation means to cover *dolus directus* in the second degree, where the volitional component of the mental element is weak — the accused does not want to cause the consequence — and it is the cognitive component that dominates — the accused was almost certain that his or her conduct would cause the proscribed consequence. However, discussions persist on whether any lower level of the mental element is captured by the wording 'will occur in the ordinary course of events'.

Most commentators argue that the default rule of Article 30 of the ICC Statute does not accommodate any standard of the mental element below the threshold of knowledge of consequences in terms of practical certainty.[43] Other voices have, instead, defended a broader conception, claiming that also some forms of conscious risk-taking with respect to consequences — that

40 T. Gal, 'Direct Commission', in L. Yanev, M.J. Ventura and M. Cupido (eds), *Modes of Liability in International Criminal Law* (Cambridge University Press, 2019) 17, at 25; Ambos and Triffterer, *supra* note 33, at 1117.

41 Finnin, *supra* note 35, at 343; M.E. Badar and S. Porro, 'Rethinking the Mental Elements in the Jurisprudence of the ICC', in C. Stahn (ed.), *The Law and Practice of the International Criminal Court* (Oxford Scholarly Authorities on International Law, 2015), at 653.

42 This wording is identical to Art. 30(3) ICCSt. concerning the gradation of 'knowledge' in relation to an element of consequence.

43 M.E. Badar and M. Bohlander, *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach* (Reprint edn., Hart Publishing, 2015), at 392; Eser, *supra* note 32, at 915; S. Porro, *Risk and Mental Element: An Analysis of National and International Law on Core Crimes* (Nomos, 2014), at 179; Finnin, *supra* note 35, at 358; K.J. Heller and M. Dubber (eds), *The Handbook of Comparative Criminal Law* (1st edn., Stanford Law Books, 2010), at 604; Bo, *supra* note 3, at 291.

under domestic criminal laws satisfy the standards of *dolus eventualis* or recklessness — would meet the requirements of Article 30 ICC Statute.[44]

The earlier practice of the ICC (the Court) offered a broad interpretation of the standard. In the *Lubanga* case, the Pre-Trial Chamber I held that Article 30 of the ICC Statute encompasses 'situations in which the suspect is aware that the risk of the objective elements of the crime may result from his or her actions or omissions and accepts such an outcome by reconciling himself or herself with it or consenting to it (also known as *dolus eventualis*)'.[45] Subsequently, however, the Court turned down this approach. At first, in the *Bemba* trial, where the ICC Pre-Trial Chamber II argued that Article 30 of the ICC Statute covers only forms of the mental element known as *dolus directus* in the first and second degree. Other forms of awareness, where the cognitive component related to the cause of consequences demands only knowledge of the mere possibility or probability, do not meet the standard of Article 30 (2)(b) and (3). The interpretation of the standard for the foreseeability of consequences that the Court brought forward is one of 'virtual certainty'.[46] This restrictive interpretation was later endorsed in the *Lubanga* appeal judgment and is, at present, the leading view in the practice of the Court.[47]

The standard established by the Court is generally understood as setting a relatively strict standard of knowledge. This stringency demands particular attention at least at two levels of the standard *dolus directus* in the second degree where the human comprehension of the consequence generated by the use of DDL systems and the process leading to that consequence may challenge the finding of the cognitive component of the suspect.

The first level concerns the fact that the cognitive component of *dolus directus* in the second degree demands the person to have certain expectations about the prohibited consequences. In this regard, commentators and the Court in its latest decision have maintained that, at a minimum, the suspect must perceive, at the time of carrying out the conduct, that it will cause the prohibited consequence unless extraordinary circumstances intervened. For instance, Finnin maintains that 'Article 30 should be interpreted to mean that the result would occur in the absence of some wholly improbable supervening event'.[48] According to the ICC Trial Chamber, this should be understood as meaning that 'it is nearly impossible for [the person] to foresee that

44 H.H. Jescheck, 'The General Principles of International Criminal Law Set Out in Nuremberg, as Mirrored in the ICC Statute', 2 *JICJ* (2004) 38, at 45; A. Gil, 'Mens Rea in Co-Perpetration and Indirect Perpetration According to Article 30 of the Rome Statute. Arguments against Punishment for Excesses Committed by the Agent or the Co-Perpetrator', 14 *International Criminal Law Review* (*ICLR*) (2014) 82, at 86 and 107; A. Cassese et al., *Cassese's International Criminal Law* (3rd edn., Oxford University Press, 2013), at 56.

45 Decision, *Lubanga* (ICC-01/04-01/06-803-tEN), Pre-Trial Chamber, 29 January 2007, § 352(ii).

46 Decision, *Bemba* (ICC-01/05-01/08-424), Pre-Trial Chamber, 15 June 2009, § 368.

47 Judgment, *Lubanga* (ICC-01/04-01/06-2842), Trial Chamber, 14 March 2012, § 1011.

48 Finnin, *supra* note 35, at 344.

the consequence will not occur'.[49] In line with this, Ambos maintains that merely anticipating or believing that the consequence is possible would not be sufficient to meet the standard under Article 30(2)(b) of the ICC Statute.[50]

The second important level follows from the former, namely that a so-called action-plan seems to be part of the cognitive component of this gradation, and thus, necessary for the proof of the mental element. Indeed, where an element of consequences in the definition of the crime in the ICC Statute is concerned, the cognitive component of *dolus directus* in the second degree is arguably not only 'to do X' but also 'to do X by doing Y, Z . . .'.[51]

Obviously, not all deviations from the suspect's wanted action-plan challenge the finding of the cognitive component in relation to the element of consequence as defined under Article 30(2)(b) (second alternative) of the ICC Statute. In the same manner, as for the level of expectations regarding the prohibited result, the suspect's action-plan does not need to foresee in all details the process leading to the consequence, but only in its essential course.[52] In other words, uncertainties about the process leading to the prohibited consequence are conceivably of concern only if they are outside the 'boundaries of what may be foreseen according to general experiences of life'.[53] This standard of 'general experience of life' seems to accurately reflect the reality in which most humans merely build solid representations of how the world works around them. Yet, they do not know every detail of how things actually work. To illustrate this point, most people probably do not know exactly how their kettle works, but they know how to use it to turn cold water into boiling water.

All in all, the ICC Statute sets relatively high standards for the mental element to exist. In fact, at a minimum, the accused needs to have a level of awareness about the proscribed consequence that his or her conduct produces, which is of 'practical certainty'. The question is then: how does the malign use of a DDL system, resulting in a war crime, or other international crime that includes a material element of consequence, challenge the finding of the mental element as defined under the ICC Statute? This question will be the focus of the following part.[54]

49 Judgment, *Katanga* (ICC-01/04-01/07-3464), Trial Chamber, 7 March 2014, § 777 (following an unofficial translation provided by Sara Porro).

50 Ambos, *supra* note 38, at 20-21.

51 E. Mayr, 'The Problem of Consequential Waywardness: Between Internalism and Externalism about Intentional Agency', in B. Kahmens and M. Stepanians (eds), *Critical Essays on 'Causation and Responsibility'* (De Gruyter, 2013) 271, at 279.

52 Guidance on the matter within the international criminal law literature is rather scarce. The findings were largely drawn from discussions in the literature on domestic criminal law. See Eser, *supra* note 32, at 918.

53 *Ibid.*

54 As highlighted earlier (see footnote 31), although the concerns related to the mental element in the two scenarios presented can largely be addressed drawing on classical criminal law doctrine of *aberratio ictus*, the parallelism that are made need to be taken with caution. This is because the use of DDL systems may raise additional challenges in establishing the causal link between the accused's conduct and the proscribed consequence.

4. The Malign Use of DDL Systems and the Level of Awareness about the Prohibited Consequences

To answer the question, we shall now turn to the two scenarios presented earlier: (a) soldier A of State Alpha wants to kill wounded soldier B of State Beta with a DDL weapon system and B dies; (b) soldier A of State Alpha wants to kill wounded soldier B of State Beta with a DDL weapon system but instead, as a result of it, wounded soldier C dies. In other words, the first scenario concerns the malign use of a DDL system resulting in the wanted consequence, while the second scenario reflects the malign use of a DDL system resulting in the same consequence but different than the one envisaged.

A. Scenario I: The Malign Use of a DDL System Resulting in the Wanted Consequence

1. Challenges

For many international law scholars, this scenario seems not to pose any major challenge to the finding of the mental element and, therefore, the ascription of individual criminal responsibility since all the ingredients are present: the human wants (*mens rea*) to commit the crime that resulted from his or her use of the DDL system (*actus reus*).[55] As Amoroso puts it, situations, where a human uses an AI system with the will to commit international crimes, make 'responsibility ascription relatively unproblematic since the agents' mental state clearly meets the *mens rea* requirement as it is generally understood in international criminal law'.[56]

This assumption likely holds for the ascription of individual criminal responsibility in scenarios like the one at hand in most domestic jurisdictions. Our present scenario essentially falls within a situation that is known in civil law countries as *dolus directus* in the first degree. As previously elucidated, in both domestic legal systems, the law related to this gradation of the mental element generally demands a very high volitional component yet only a very low cognitive component — i.e. it does not require that the accused be certain that he or she will succeed in bringing about the consequence. In other words, the mental element could be established even if the accused is not 100% sure that he or she will succeed in bringing about the consequence, so long as it was his or her will to produce that consequence. Therefore, in our scenario where the accused clearly wanted to produce the prohibited consequence by

55 See for instance: D. Amoroso and B. Giordano, 'Who Is to Blame for Autonomous Weapons Systems' Misdoings?' in E. Carpanelli and N. Lazzerini (eds), *Use and Misuse of New Technologies: Contemporary Challenges in International and European Law* (Springer International Publishing, 2019) 211, at 217; Bo, *supra* note 3; A.G. Jain, 'Autonomous Cyber Capabilities and Individual Criminal Responsibility for War Crimes', in R. Liivoja and A. Väljataga (eds), *Autonomous Cyber Capabilities and International Law* (NATO Cooperative Cyber Defence Centre of Excellence, 2021), at 7.

56 Amoroso and Giordano, *supra* note 55, at 217.

using a DDL system, his or her lack of 100% certainty about the prohibited consequence being produced by this use would not affect the finding of the mental element and therefore the ascription of individual criminal responsibility. Thus far, this seems to be the dominant view within criminal law scholarship and may be the reason why individual criminal responsibility concerns regarding the malign use of DDL systems have remained largely unaddressed.

Yet, what if one accepts the view that *dolus directus* in the first degree as covered under Article 30 of the ICC Statute sets a higher standard than under domestic law — meaning requiring not only a high volitional component but also a high level of awareness about the consequences?

Following this view, we would likely come to a different conclusion. We would be left with the undesired result that, even where the human wants to use a DDL system to commit an international crime, the finding of the mental element and, consequently, the ascription of individual criminal responsibility may be challenged. Unless it can be shown that he or she knew that 'in the ordinary course of events' the use of the DDL system will generate the wanted consequence, he or she would not satisfy this gradation of the mental element.

This is because, as presented earlier, due to the inherent lack of predictability and explainability of DDL systems, the operators' capacity to anticipate the systems' consequences, and the manner with which it generates the consequence, is often significantly reduced. This does not mean that it is impossible for an operator to anticipate the DDL system's consequence and the way it produces the consequence with the level of certainty required under Article 30 of the ICC Statute. But, it would certainly require a certain degree of predictability and explainability of the DDL system that allows the operator to have a reliable understanding of *what the system will do* and *why it does what it does*. Still, ensuring such a level of predictability remains extremely difficult — if not impossible — for systems that are developed to operate in dynamic, complex and highly fluid environments, such as armed conflicts.[57]

In summary, the finding of the mental element in the scenario at hand does not pose major responsibility challenges, unless a stringent standard of knowledge about the consequence element of the crime is required. In other words, the malign use of a DDL system for the commission of international crimes generates responsibility gaps essentially when the mental element requires a high cognitive component. This is because present DDL systems lack explainability and predictability, which in turn significantly reduce the human ability to anticipate the system's effects and comprehend why the system produced those effects. The following subsection, therefore, explores ways to address the responsibility gaps generated by the stringent standard of awareness under Article 30 of the ICC Statute.

57 See discussion in Part 2 A on Human Comprehension of DDL Systems' Use: Predictability and Explainability Challenges.

2. Possible Solutions

One way to address the risk of a responsibility gap would entail the lowering of the stringent standard of knowledge under Article 30 of the ICC Statute by recognizing its aptitude to incriminate risk-taking behaviours, such as *dolus eventualis* existing in civil law systems. Within the responsibility gap literature and beyond there has been an important scholarly debate on the question of whether the language 'will occur in the ordinary course of events' in Article 30(2)(b) (and its identical wording in Article 30(3)) of the ICC Statute covers situations of *dolus eventualis*.[58]

For the present discussion, it is not necessary to enter into the details of the debate. It shall focus instead on assessing the value of *dolus eventualis* as a means of reducing the responsibility gap in the situation of malign use of a DDL system. Ultimately, this may provide new elements to consider within the broader discussion about the inclusion/exclusion of *dolus eventualis* in the legal standard of Article 30 of the ICC Statute.[59] Indeed, this gradation offers a lower standard of the mental element to be applied, not only to operators of DDL systems but also to any accused of an international crime under the ICC Statute, considering the fundamental principles of legality and equality.

Within the literature on the responsibility gap, scholars have indeed recognized its value for better embracing the current development of intelligent systems.[60] *Dolus eventualis* allows for the attribution of responsibility for international crimes where a human using a DDL system envisages and accepts the risk that a crime will be committed. Such a gradation of the mental element demands more care and attention when DDL systems are used. Eventually, it would push users and developers to adopt clear standards around the use of such systems, clarifying the level of standard of care that needs to be fulfilled so that risks of the use of such systems can be excluded. Considering the development of increasingly more complex DDL systems, it would be crucial for such standards to address questions related to the level of explainability and predictability necessary when such systems are used. All these aspects considered, *dolus eventualis* could, above all, reduce the possibility for perpetrators to hide behind the unpredictability and lack of explainability of these systems in situations such as the present one.

58 Badar and Porro, *supra* note 41, at 654–662; Finnin, *supra* note 35, at 333–336; Eser, *supra* note 32, at 932.

59 On the broader discussion regarding *dolus eventualis* and Art. 30 ICCSt. see for instance: M.E. Badar, 'Dolus Eventualis and the Rome Statute Without It?' 12 *New Criminal Law Review: An International and Interdisciplinary Journal* (2009) 433–467; A. Cassese, P. Gaeta and J.R.W.D. Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary* (Oxford University Press, 2002); K. Ambos, 'Critical Issues in the *Bemba* Confirmation Decision', 22 *Leiden Journal of International Law* (2009) 715; Finnin, *supra* note 35.

60 See for instance: Bo, *supra* note 3; A. Seixas-Nunes, 'Accountability and Liability for the Deployment of Autonomous Weapon Systems', in idem (ed.), *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective* (Cambridge University Press, 2022) chapter 5, at 218–220.

B. Scenario II: The Malign Use of DDL Systems Resulting in the Same Consequence but Different than the One Intended

1. Challenges

This second scenario could potentially raise difficulties for the ascription of criminal responsibility because the prohibited consequence that the person wanted (soldier A) through using the DDL system occurs on a person (wounded soldier C) other than the one envisioned (wounded soldier B). In this case, the nature of the crime is the same as the one wanted (killing a person *hors de combat*), but the victim is not the one intended by soldier A (that is, wounded soldier C instead of B).[61]

The added difficulty for ascribing individual criminal responsibility for the crime committed in this scenario arises from the fact that the suspect may not only lack knowledge about the consequence element of the crime but also the intent to harm the envisaged person. In other words, the deviation from the wanted proscribed consequence renders more difficult the proof of the two components required to prove the mental element under Article 30 of the ICC Statute. Consequently, the ascription of individual criminal responsibility may be challenged, creating an additional responsibility gap in situations of malign use of DDL systems for the commission of international crimes.

2. Possible Solutions

At first glance, the closest avenue available in the present scenario is to prosecute the accused (soldier A) for the attempt of committing a war crime against wounded soldier B, drawing on Article 25(3)(f) of the ICC Statute.[62] This definition offers some information on the objective elements of the attempt[63] which are: (i) the commencement of the execution of an international crime by means of essential steps and (ii) the non-occurrence of this crime for circumstances independent of the person's will. However, it remains silent

61 It is important to distinguish the present scenario from the situation where the deviations in the prohibited consequence are irrelevant because the human is completely indifferent to the harm his or her conduct may cause. For instance, a person uses a DDL system to fire at a group of people without aiming at a specific victim. In this case, it does not matter which individuals are harmed since the will exists in relation to all the victims. On this point see also: Eser, *supra* note 32, at 918.

62 It defines '[a]ttempts to commit such a crime by taking action that commences its execution by means of a substantial step, but the crime does not occur because of circumstances independent of the person's intentions. However, a person who abandons the effort to commit the crime or otherwise prevents the completion of the crime shall not be liable for punishment under this Statute for the attempt to commit that crime if that person completely and voluntarily gave up the criminal purpose'.

63 For the present discussion, it will be assumed that objective elements are established. Nevertheless, it is important to note that the jurisprudence of international criminal courts and tribunals on attempt is rather scarce having left open some delicate questions of interpretation regarding the objective elements. On this point see: J. de Hemptinne, 'Attempt', in Cupido, Ventura, and Yanev (eds), *supra* note 40, chapter 11.

about the required mental element. The dominant view is that the accused must, in principle, share the same intention as an accused of a completed offence. This approach was adopted by the ICC Pre-Trial Chamber in the *Katanga* case.[64] It is shared by most commentators[65] and seems to be reflected in many Western domestic legal systems.[66] In the context of the ICC Statute for crimes with an element of consequence, this entails that the accused must meet the standard of *dolus directus* in the first or second degree. As demonstrated in relation to the first scenario, it is essentially the cognitive component that would be difficult to prove because present DDL systems significantly reduce the human ability to anticipate the system's effects and comprehend why the system produced those effects. Thus, unless the stringent standard of knowledge about the consequence element of the crime foreseen under Article 30 of the ICC Statute is required — i.e. in situations of *dolus directus* in the second degree — there would be no major challenge in establishing the mental element. Following the majority's understanding of *dolus directus* in the first degree, soldier A would therefore be criminally responsible for the attempted killing of wounded soldier B.

This being said, even the successful ascription of responsibility for an attempted war crime against soldier B would leave soldier A's responsibility regarding the killing of wounded soldier C unaddressed. The most apparent difficulty here is that soldier A lacks the mental element since he or she had neither the will to cause the death of wounded soldier C nor the awareness that his or her use of the DDL system will cause that consequence. Hence, it would be impossible to establish *dolus directus* in the first or second degree in relation to the killing of wounded soldier C. But soldier A could, at best, be held responsible for negligence or lack of due care[67] as it is widely known that the mental element under the ICC Statute excludes the possibility of ascribing individual criminal responsibility for negligence.[68] Accordingly, we would be left with an important responsibility gap regarding the victim actually harmed (wounded soldier C) by the use of a DDL system.

Another avenue that may be found in domestic legal systems to address situations of deviations in the prohibited consequence as in our second scenario, is through the doctrine known as *aberratio ictus* (accidental harm to a person) in civil law or 'transferred intent' in common law systems. Indeed, these terms are used to designate situations in which the accused effectively

64 Decision, *Katanga and Ngudjolo Chui* (ICC-01/04-01/07-717), Pre-Trial Chamber, 30 September 2008, § 460.

65 K. Ambos, *Treatise on International Criminal Law: Volume I: Foundations and General Part* (2nd edn., Oxford University Press, 2021), at 243–244; Cassese, Gaeta and Jones (eds), *supra* note 59, at 811.

66 On this point see: de Hemptinne, *supra* note 63, at 347.

67 A culpable lack of due care arises when a 'person fails to exercise the care that is incumbent on him in the circumstances and commensurate with his personal capabilities'. See for instance Art. 12 of the Swiss Criminal Code of 21 December 1937 (Status as of 23 January 2023).

68 Badar and Porro, *supra* note 41, at 664. See: Judgment, *Lubanga* (ICC-01/04-01/06-A-5), Appeals Chamber, 1 December 2014, § 438; Decision, *Bemba* (ICC-01/05-01/08-424), Pre-Trial Chamber, 15 June 2009, § 360.

directs his or her conduct against a specific person (or object), but fails to harm it, the harmful consequence being produced in another person (or object).[69]

The present scenario of *aberratio ictus* needs to be distinguished from an e*rror in persona vel objecto* (mistaken the identity of a person). The difference between the two lies in the fact that in the latter, the accused killed the person he or she had individualized as a target, while in the case of *aberratio ictus*, the accused does not kill the person he or she had individualized.[70] Thus, e*rror in persona vel objecto* is an error as to the identity of the victim or object — i.e. A kills X believing he or she is Y. Most importantly, it would not alter the intent since there is no relevant deviation from the actual causal process that could impair the proper formation of the will — at least where the mistaken persons or objects are of equal value. Accordingly, in the case of an error, domestic jurisdictions generally agree that the accused (A) would be held criminally liable for the intentional killing (of Y).[71]

Yet, regarding *aberratio ictus* differing views exist on the ascription of criminal responsibility, especially in situations where the objects or persons concerned are of equal value. Without exploring them in depth, the existing views on the doctrine can be regrouped in the following two main approaches. The first approach defends that the intent can be transferred to the victim killed even if it is not the wanted victim. The most common justification for this view is that the object or person being harmed should not matter as long as it is of the same kind and equal value under the law — in our scenario, a human being. This idea has been expressed in phrases such as 'the perpetrator wanted to kill a human being and killed a human being'.[72] Following this line of reasoning, a situation of *aberratio ictus* would not dispense the accused from being responsible for intentional killing. This is the dominant approach in English[73] and Italian law.[74] A different approach is provided by the German prevalent academic opinion and case law,[75] maintaining that in situations of *aberratio ictus* the accused would be 'held liable for attempted killing in relation to the intended victim and for negligent manslaughter with regard to the actual victim'.[76] For proponents of this view, the accused cannot be held liable for intentional killing, insisting on the need to prove intent regarding the object or person the accused has individualized.

As within most domestic jurisdictions, the ICC Statute does not explicitly regulate such situations of deviant crimes. Therefore, drawing on Article 21 of its Statute, the Court would need to demonstrate that the doctrine of *aberration*

69 J.M. Silva-Sanchez, 'Aberratio ictus und objektive Zurechnung', 101 *Zeitschrift für die gesamte Strafrechtswissenschaft* (1989) 352, at 352.

70 M.E. Badar, 'Mens Rea - Mistake of Law & Mistake of Fact in German Criminal Law: A Survey for International Criminal Tribunals', 5 *ICLR* (2005) 203, at 239.

71 *Ibid.*, at 218; Silva-Sanchez, *supra* note 69, at 353; Eser, *supra* note 32, at 938.

72 Silva-Sanchez, *supra* note 69, at 357.

73 S. Eldar, 'The Limits of Transferred Malice', 32 *Oxford Journal of Legal Studies* (2012) 633, at 634.

74 Art. 82(1) of the Italian Criminal Code.

75 Silva-Sanchez, *supra* note 69, at 355.

76 Eser, *supra* note 32, at 918.

ictus represents a general principle of law derived from domestic legal systems. Still, the Court's interpretative power is not unlimited. In line with the principle *nullum crimen sine lege* foreseen under Article 22 of its Statute, the Court is bound to a strict interpretation of the text, that, in case of ambiguity, should be interpreted in the accused's favour. This being said, we can for the time being only speculated about the approach of the *aberratio ictus* doctrine the Court may ultimately choose to adopt.

An appropriate starting point is the preeminent role that intentional conduct occupies in the ICC Statute's default definition of the mental element. Accordingly, the second approach giving more weight to the concretization of intent would appear more compliant with the ICC Statute's stringent standard of the mental element and, consequently, also with the principle of legality and the principle of strict construction.[77] Nevertheless, this would leave us with the same problem raised earlier concerning the challenges for the ascription of responsibility regarding the actual victim of the use of the DDL system (wounded soldier C). In other words, adopting the second approach would mean that the accused could be held responsible, at the utmost, for the attempted killing of the wanted victim (wounded soldier B). As a result, we would be left with a significant responsibility gap, at least, with respect to the victim actually harmed by the use of a DLL system (wounded soldier C). Some might be inclined to suggest that this undesirable consequence should be a reason for the Court to adopt the first position.[78]

Whichever road the Court chooses to take, to address such situations of malign use of DDL systems will likely necessitate a compromise solution, requiring the balancing of two core objectives of criminal justice: on the one hand, the right of victims to be compensated for the harm suffered by not allowing perpetrators to hide their culpability behind potentially unpredictable and insufficiently explainable DDL systems; and, on the other hand, avoiding the introduction of a standard of a mental element that would stretch the *nullum crime sine lege* principle in a manner that would undermine the right of the accused.

5. Conclusion

This article explored a new dimension of the responsibility gap literature by examining the impact the malign use of DDL systems has on the finding of the mental element as defined under Article 30 of the ICC Statute, and thus, the ascription of individual criminal responsibility. Drawing on two fictive scenarios, it demonstrated that the malign use of DDL systems creates new criminal responsibility concerns that require different legal responses.

Most importantly, it became apparent that the adoption of some of these responses may confront our societies with difficult criminal policy choices on

77 On this point see also: Fattori, *supra* note 8.
78 Eser, *supra* note 32, at 918.

conflicting values: do we wish to enhance the protection of our societies from potentially harmful technological developments? Or do we want to uphold some of our fundamental principles, such as the *nullum crime sine lege* principle that is central to the protection of any accused? As with most novel phenomena creating new challenges, we need to determine the extent to which we are ready to sacrifice certain fundamental values of our present societies in order to sustain others.

To conclude, this article does not offer an exhaustive list of criminal responsibility challenges raised by the malign use of DDL systems. Likewise, it does not aim at providing any conclusive answers to the new responsibility gap issues identified. Instead, the present discussion hopes to offer a starting point for the continuation of this conversation that, in light of the current developments, is crucial and urgently needed.