The use of the 'Lavender' in Gaza and the law of targeting: AI-decision support systems and facial recognition technology

Andersin, E.M.A.

BRILL
NIJHOFF

JOURNAL OF INTERNATIONAL HUMANITARIAN LEGAL STUDIES (2025) 1–35

INTERNATIONAL HUMANITARIAN LEGAL STUDIES

brill.com/ihls

The Use of the 'Lavender' in Gaza and the Law of Targeting: AI-Decision Support Systems and Facial Recognition Technology

Emelie Andersin | ORCID: 0009-0008-3886-832X
PhD Fellow, Leiden University College, Leiden University, The Hague, The Netherlands
e.m.a.andersin@luc.leidenuniv.nl

Received 25 September 2024 | Accepted 28 April 2025 | Published online 23 May 2025

Abstract

On 7 October 2023, the long-lasting Israel-Hamas conflict escalated significantly in scale and violence. Reports reveal that the Israeli military have employed Artificial Intelligence Decision Support Systems ('AI-DSS') to identify individuals in targeting situations. Another report has indicated that the Israeli military used facial recognition technology ('FRT') to identify Palestinians in Gaza. Scholars are debating the legality of AI-DSS under International Humanitarian Law ('IHL'), and the extent to which military commanders can rely on AI for targeting decisions. This article describes the challenges in human-machine interaction with a focus on algorithmically generated recommendations and the responsibility of military commanders in this regard. The article concludes that, while the use of FRT can enhance accuracy in identifying individuals and support adherence to IHL obligations, its effectiveness depends on the operational environment. It also emphasises the importance of improving military commanders' technical literacy of AI-DSS and ensuring that sufficient time is taken to verify the accuracy of algorithmically generated targets.

Keywords

Gaza – AI-DSS – facial recognition technology – artificial intelligence – bias – opacity – human-machine interaction – military commander – international humanitarian law

Published with license by Koninklijke Brill BV | DOI:10.1163/18781527-BJA10119
 | ISSN: 1878-1373 (print) 1878-1527 (online)

1 Introduction

Modern warfare has become a technological battle. Militaries are required to deal with an overwhelming flow of information, complexity, and time pressure. For militaries to be effective, they must evaluate enormous amounts of data, especially in complex and high-intensity environments.[1] To ensure effectiveness and make faster decisions, States have a growing interest in developing artificial intelligence ('AI') tools that can more quickly process these large quantities of data.[2] Some militaries, such as the United States ('US'), are developing AI-based systems that generate recommendations to assist military commanders with making effective decisions.[3] AI-decision support systems ('AI-DSS') are 'computerized tools that are designed to aid humans in making complex decisions by presenting information that is relevant'[4] for the tasks they are designed to perform. They assist militaries in collecting data, predicting patterns, or making recommendations based on large quantities of information.[5] At the same time, there is a growing trend of States using facial recognition technology ('FRT')[6] in wartime.[7] States are pursuing this technology because FRT can be used as input for AI-based systems to *identify* human targets based on their visual characteristics.[8] Facial recognition can improve accuracy and effectiveness to identify or verify known enemies'

1 Merel A C Ekelhof, 'Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting' (2018) 71 Naval War College Review 61, 76.

2 Ashley Deeks, 'Coding The Law of Armed Conflict: First Steps' in Matthew C Waxman and Thomas W Oakley (eds), *The Future of Armed Conflict* (Oxford University Press 2022), 45.

3 Sydney Freedberg, 'ATLAS: Killer Robot? No. Virtual Crewman? Yes.' (*Breaking Defense*, 4 March 2019) <https://breakingdefense.com/2019/03/atlas-killer-robot-no-virtual-crewman-yes/>; Dustin Lewis, Naz Modirzadeh, and Gabriella Blum, 'The Pentagon's New Algorithmic-Warfare Team' (*Lawfare*, 26 June 2017) <https://www.lawfaremedia.org/article/pentagons-new-algorithmic-warfare-team>.

4 Arthur Holland Michel, 'Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making' (May 2024) ICRC Observations on External Report, 13.

5 Nehal Bhuta, Susanne Beck, and Robin Geiβ, 'Present Futures: Concluding Reflections and Open Questions on Autonomous Weapons Systems' in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016), 347–383.

6 Anil K Jain, Arun Ross, and Salil Prabhakar, 'An Introduction to Biometric Recognition' (2004) 14 IEEE Transactions on Circuits and Systems for Video Technology.

7 Alison Mitchell, 'Distinguishing Friend from Foe: Law and Policy in the Age of Battlefield Biometrics' (2012) 50 Canadian Yearbook of International Law 289; William C Buhrow, *Biometrics in Support of Military Operations: Lessons from the Battlefield* (CRS Press 2017).

8 William H Boothby, 'Biometrics' in William H Boothby (ed), *New Technologies and the Law in War and Peace* (Cambridge University Press 2021), 397.

identities at a distance.[9] These AI-tools and FRT have been developed to make effective and accurate decisions in military targeting situations.

On 7 October 2023, the Israel-Hamas conflict escalated after Hamas and other armed groups attacked the southern part of Israel. This attack involved the killing and hostage-taking of civilians, and sexual violence against civilians.[10] In response, Israel launched a military operation called 'Operation Iron Swords' in the Gaza Strip. The +972 Magazine and the Local Call[11] reported that the Israeli Defense Forces ('IDF') used an AI-DSS in Gaza, known as 'the Lavender'.[12] Interviews with IDF intelligence officers revealed that this system was designed by Israel to identify Palestinians who might have links to Hamas and other armed groups, such as the Palestinian Islamic Jihad ('PIJ'), as potential targets for strikes. These recommendations are sent to Israeli intelligence analysts where, it is reported, they review targets and sometimes send reviewed recommendations to military commanders.[13] The final decision to approve attacks against targets relies upon the military commander.[14] Yet it is not known exactly what data and intelligence input Lavender uses to make recommendations. According to the IDF, recommendations come from a wide variety of sources, namely, geospatial intelligence, signal intelligence, human sources, and open-source information.[15] On 24 March 2024, the New York Times reported that the Israeli military has established a mass surveillance program based on facial recognition in Gaza that identifies individuals without their consent or knowledge. It uses facial recognition to recognise an individual through cameras, military checkpoints, and from drone footage.[16] In 2023, a commander of the Unit 8200 in the IDF explained that they are able to locate 'dangerous' people based on input from a list of individuals who have been entered into the system.[17] This seems to suggest that the IDF may be using FRT

9 Leah West, 'Face Value: Precaution versus Privacy in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE Publications 2022), 136.

10 Abdelali Ragad et al, 'How Hamas Built a Force to Attack Israel on 7 October' (*BBC*, 27 November 2023) <https://www.bbc.com/news/world-middle-east-67480680>.

11 The +972 Magazine and the Local Call are independent and non-profit magazines.

12 Yuval Abraham, '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza' (*+972 Magazine*, 3 April 2024) <https://www.972mag.com/lavender-ai-israeli-army-gaza/>.

13 ibid.

14 IDF Website, 'The IDF's Use of Data Technologies in Intelligence Processing'(*Israel Defense Forces*, 18 June 2024) <https://www.idf.il/210062>.

15 ibid.

16 Sheera Frenkel, 'Israel Deploys Expansive Facial Recognition Program in Gaza' (*The New York Times*, 27 March 2024) <https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html>.

17 רורט ידעי לש רתוי ריהמ יוליגו גוויס תרשפאמ תיתוכאלמ הניב :8200, תיתוכאלמה הניבה זכרמ דקפמ [Commander of the Artificial Intelligence Center, 8200: Artificial Intelligence

as input to identify previously known individuals from enrolled biometrics in the database.[18]

According to +972 Magazine and Local Call, the Lavender generated at least 37,000 target recommendations during the first six weeks of the conflict. These included both high and low-ranking operatives.[19] The +972 Magazine and Local Call report indicates that the system 'in general' was reported to have 90 percent accuracy, where the IDF sometimes authorised an airstrike. The IDF relied on another tracking system called 'Where's Daddy?'. This system was used to track suspected militants, and signalled to the Israeli military when they entered their home. In some cases, the IDF would mark the house for bombing while the target's family members were present in the home.[20]

Another AI-system used by the IDF is called 'Fire Factory' and relies on 'data about military-approved targets to calculate munition loads, prioritize and assign thousands of targets to aircraft and drones, and propose a schedule'.[21] According to Tal Mimran and Gal Dahan, this AI-system is used for different tasks, such as analysing data about previous targets and 'the prioritization and allocation of targets'.[22] The Lavender is also used in conjunction with another AI-DSS, known as 'the Gospel', which marks buildings and structures as targets that alleged militants operate from. These AI-based systems are operated by the IDF's elite intelligence Unit 8200.[23] Importantly, the reporting about the IDF's use of the Lavender remains highly limited and it is difficult to verify the information from these reports. Yet, given the unprecedented civilian harm and destruction in Gaza, it is pivotal to address such reporting.

Enables Faster Classification and Detection of Terrorist Targets] (*Israel Defense*, 14 February 2023) <https://www.israeldefense.co.il/node/57256>.

18 For more details regarding previous collection, see Privacy International, 'Biometrics and Counter-Terrorism:
Case Study of Israel/Palestine' (May 2021) <https://privacyinternational.org/sites/default/files/2021-06/PI%20Counterterrorism%20and%20Biometrics%20Report%20Israel_Palestine%20v7.pdf>, 9.

19 Abraham (n 12).

20 ibid.

21 Marissa Newman, 'Israel Quietly Embeds AI Systems in Deadly Military Operations' (*Bloomberg*, 16 July 2023) <https://www.bloomberg.com/news/articles/2023-07-16/israel-using-ai-systems-to-plan-deadly-militaryoperations?embedded-checkout=true&leadSource=uverify%20wall>.

22 Tal Mimran and Gal Dahan, 'Artificial Intelligence in the Battlefield: A Perspective from Israel' (*Opinio Juris*, 20 April 2024) <https://opiniojuris.org/2024/04/20/artificial-intelligence-in-the-battlefield-a-perspective-fromisrael/>.

23 Yuval Abraham, "A Mass Assassination Factory': Inside Israel's Calculated Bombing of Gaza' (*+972 Magazine*, 30 November 2023) <https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/>.

Scholars and experts have expressed deep concern for the massive civilian harm inflicted on the Gazan population.[24] At the time of writing, more than 40,000 Palestinians have been killed in Gaza since 7 October 2023, at least 92,401 Palestinians have been wounded, and more than half of Gaza's buildings destroyed or damaged.[25] This scale of civilian casualties and damage to civilian infrastructure raises serious concern about the use of AI in military targeting decisions and its ability to mitigate civilian harm in battlefield targeting.

The purpose of this paper is to unpack the legal challenges arising from military commanders relying on AI-DSS in targeting situations. Moreover, it aims to explore the role of human-machine interaction in this context, specifically the use of algorithmically generated 'targets' by military commanders. It will examine the application of the law of targeting, using the Lavender as a case study. The use of the Lavender as a case study will contribute to a broader understanding of how international humanitarian law ('IHL') applies to the use of such systems. Thus, this paper aims to highlight the general existence and use of AI-DSS since this technology is likely to become more prevalent and relevant in future armed conflicts.

This paper will focus on AI-DSS receiving input from FRT because this is an underexplored area in IHL. By addressing this gap, this article aims to highlight and examine how AI-DSS in combination with FRT may trigger legal concerns. The inaccuracy of the FRT and the ability to process vast amounts of data by AI-DSS in armed conflicts can raise unforeseen consequences. It will not examine the applicability of other areas of law that are potentially relevant, such as International Human Rights Law ('IHRL')[26] and Data Protection Law ('DPL').[27] It does not dismiss the relevance and importance of examining these legal areas, but recognises that such topics are outside the scope of this paper.

24 University Network for Human Rights et al, 'Genocide in Gaza: Analysis of International Law and Its Application to Israel's Military Actions Since October 7, 2023' (15 May 2024) <https://static1.squarespace.com/static/66a134337e960f229da81434/t/66fb05bb0497da4726e125d8/1727727037094/Genocide+in+Gaza+-+Final+version+051524.pdf>.

25 Julia Frankel, 'With Gaza's Death Toll over 40,000, Here's the Conflict by Numbers' (*The Associated Press*, 15 August 2024) <https://apnews.com/article/israel-hamas-gaza-war-palestinians-statistics-400007ebec13101f6d08fe10cedbf5e172dde>.

26 West (n 9).

27 Asaf Lubin, 'The Rights to Privacy and Data Protection under International Humanitarian Law and Human Rights Law' in Robert Kolb, Gloria Gaggioli, and Pavle Kilibarda (eds), *Research Handbook on Human Rights and Humanitarian Law: Further Reflections and Perspectives* (Edward Elgar Publishing 2022).

2 The Use of Technology in Armed Conflicts

This section aims to explain the technologies that are relied upon to identify individuals' identity through FRT and generate recommendations for targeting. Firstly, it will briefly explain what FRT is, its use in the civilian domain, and in armed conflict. Secondly, it will discuss what AI-DSS are, their functions, and their military use.

2.1 *The Use of FRT to Identify or Verify Individual Identity*

Each individual has unique features because of their facial characteristics.[28] Facial recognition is a biometric modality[29] that aims to identify or verify an individual's identity through automated recognition, based on their facial characteristics.[30] Facial recognition systems are usually AI-powered and rely on algorithms and machine learning to detect, process, and recognise individuals. The process 'treats the face as an index of identity'[31] in the collection of a face by utilising an algorithm.[32] It maps out face patterns that are converted into a mathematical representation, and compared against previously enrolled biometrics in a database to find the identity of that person.[33] The purpose of facial recognition is threefold: (1) identification, (2) verification, and (3) classification of an individual's identity. In the identification process, a biometric system runs a sample against all previously collected data and conducts a one-to-many recognition process to *identify* an unknown person. In the verification process, the system conducts a one-to-one comparison when an individual has claimed an identity to *verify* if that person is who they claim to be.[34] In the classification process, the system extracts information based on an individual's characteristics to *classify* an individual's emotions,[35] gender, or race.[36]

28 Marcus Smith and Seumas Miller, *Biometric Identification, Law and Ethics* (Springbriefs in Ethics 2021), 22.

29 Biometrics Institute, 'Types of Biometrics' <https://www.biometricsinstitute.org/what-is-biometrics/types-ofbiometrics/> accessed 17 January 2024.

30 Jain, Ross and Prabhakar (n 6).

31 Smith and Miller (n 28), 22–23.

32 Kelly A Gates, *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance* (New York University Press 2011), 8.

33 For more details, see Gates (n 32), 8; Lucas Introna and Helen Nissenbaum, 'Facial Recognition Technology: A Survey of Policy and Implementation Issues' (2009) Center for Catastrophe Preparedness and Response, 15–16.

34 Mitchell (n 7).

35 Joy Buolamwini et al, 'Facial Recognition Technologies: A Primer' (29 May 2020) <https://globaluploads.webflow.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14_FRTsPrimerMay2020.pdf>.

36 European Union Agency for Fundamental Rights, 'Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement' (2019)

FRT has binary outcomes: the result gives either a positive match (there is a likelihood that the two templates belong to the same person) or a negative match (there is a likelihood that the two templates do not belong to the same person). Yet the final result is technically never a 'yes' or 'no', but instead a *probability* because an 'algorithm never returns a definitive result, but only probabilities'.[37] Thus, FRT is measured on a confidence score of whether two different templates belong to the same person. Higher confidence scores indicate that there is a likely positive match.[38]

There are two different rates of errors: (i) false positives (type I error) and (ii) false negatives (type II error). False positives are when the system generates a 'positive' match of a person's face enrolled in a biometric database, but the match is incorrect. False negatives are when the systems fail to generate a match between a person's face to an image contained in a database, whereas that person is actually enrolled in the database.[39] The necessary threshold of confidence scores depends on several factors: whether the environment is controlled, whether a human is supervising the facial recognition system, and the sensitivity of that environment. However, if there is a higher confidence threshold to avoid false positives, it allows for more false negatives, and if it has a lower confidence threshold it allows for more false positives.[40]

FRT has increasingly become a part of the security domain. It identifies suspects in public spaces, usually undertaken by law enforcement. These technologies can be used for counter-terrorism purposes, by comparing footage obtained from closed-circuit television ('CCTV') footage cameras against databases of facial images from (for example) a watchlist.[41] For instance, the Israeli authorities monitor civilians living in the Occupied Palestinian Territory ('OPT') with FRT.[42] At the same time, the use of facial recognition plays a larger

≤https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper.pdf> accessed 7 January 2024.

37 ibid, 9.

38 European Data Protection Board, 'Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement' (Version 2.0, adopted on 26 April 2023).

39 William Crumpler, 'How Accurate are Facial Recognition Systems – and Why Does It Matter?' (*Center for Strategic & International Studies*, 14 April 2020) <https://www.csis.org/blogs/strategic-technologies-blog/how-accurateare-facial-recognition-systems-and-why-does-it>.

40 ibid.

41 Information Commissioner's Opinion, 'The Use of Live Facial Recognition Technology in Public Places' (18 June 2021) < https://ico.org.uk/media/2619985/ico-opinion-the-use-of-lfr-in-public-places-20210618.pdf>.

42 For more extensive details, see Rohan Talbot, 'Automating Occupation: International Humanitarian and Human Rights Law Implications of the Deployment of Facial Recognition Technologies in the Occupied Palestinian Territory' (2020) 102 International

role and has become more ubiquitous in armed conflicts.[43] The US has been a leading actor by introducing the use of biometrics in armed conflicts following the 2003 invasion of Iraq.[44] Between 2008 and 2017, it used biometrics to capture or kill 1,700 individuals, and denied access to 92,000 individuals from military facilities using biometric data.[45] The Israeli authorities relied on FRT to identify the dead after the 7 October attack.[46] Further, the Ukrainian military used FRT in the Russia-Ukraine war to identify the dead from their own casualties and Russian casualties.[47]

2.2 *The Use of AI-DSS to Find, Select, and Recommend Targets*

Generally speaking, AI-DSS are based on either deterministic models or non-deterministic models.[48] Traditional DSS are based on deterministic computer models that produce the same output, are predictable, and do not involve randomness.[49] While these systems are predictable and will always produce constant outcomes, they are limited in their ability to process more complex issues and realities.[50] By contrast, non-deterministic models, also known as 'stochastic' models, produce outcomes that involve levels of unpredictability

Review of the Red Cross 823; Omar Yousef Shehabi, 'Emerging Technologies, Digital Privacy, and Data Protection in Military Occupation' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE Publications 2022).

43 Marten Zwanenburg, 'Know Thy Enemy: The Use of Biometrics in Military Operations and International Humanitarian Law' (2021) 97 International Law Studies 1404.

44 Spencer Ackerman, 'U.S. Holds On to Biometrics Database of 3 Million Iraqis' (*Wired*, 21 December 2011) <https://www.wired.com/2011/12/iraq-biometrics-database/>; Annie Jacobsen, *First Platoon: A Story of Modern War in the Age of Identity Dominance* (Penguin 2021).

45 United States Government Accountability Office, 'DOD Biometrics and Forensics: Progress Made in Establishing Long-Term Deployable Capabilities, but Further Actions are Needed' (August 2017) <https://www.gao.gov/assets/gao-17-580.pdf> accessed 2 February 2024.

46 Masha Borak, 'Israel is Using Amazon Rekognition to Locate Missing and Dead' (*Biometric Update*, 23 October 2023) <https://www.biometricupdate.com/202310/israel-is-using-amazon-rekognition-to-locate-missing-anddead>.

47 Drew Harwell, 'Ukraine is Scanning Faces of Dead Russians, then Contacting the Mothers' (*Washington Post*, 15 April 2022) <https://www.washingtonpost.com/technology/2022/04/15/ukraine-facial-recognition-warfare/>.

48 Holland Michel (n 4),18.

49 Agnieszka Lazarowska, 'A New Deterministic Approach in a Decision Support System for Ship's Trajectory Planning' (2017) 71 Expert Systems with Applications 469; Jorge Vargas Florez et al, 'A Decision Support System for Robust Humanitarian Facility Location' (2015) 46 Engineering Applications of Artificial Intelligence 326.

50 Priya Narayanan et al, 'First Year Report of ARL Director's Strategic Initiative (FY20–23): Artificial Intelligence (AI) for Command and Control (C2) of Multi-Domain Operations (MDO)' (DEVCOM Army Research Laboratory, May 2021), 2.

and randomness. These models are trained on datasets by feeding 'input' with examples of desired outputs.[51] Therefore, they are not coded on constrained values, with the purpose of capturing more complex realities. This is useful in the military domain, because the military battlefield is not always predictable and requires taking into account real-world intricacies.[52] With introducing *uncertainties* in the lack of exact knowledge for its outcomes, it may not be possible to understand why the model has generated a certain output, repetitive.

With the help of machine learning, the programmer trains the system to perform the algorithm's tasks and learns while providing recommendations.[53] It uses data to identify patterns and characteristics to produce outcomes based on the input data.[54] This can improve the speed of decision-making and detect patterns in large quantities of data.[55] An algorithm is 'a sequence of computational steps that transform the *input* into the output' (emphasis added).[56] This process can take form under (1) *supervised* or (2) *unsupervised learning*. The former is when a developer teaches the algorithms under supervision by labelling objects and provides feedback in either classification or regression. By contrast, unsupervised training occurs when algorithms learn by themselves to discover patterns without supervision and cluster unlabelled data in the provided data.[57] For example, if a person shares sufficient characteristics with an individual identified as a civilian directly participating in the hostilities ('DPH'), the system could label that individual as DPH. Because non-deterministic models are based on the likelihood or probability that the individual has DPH status, it is not based on absolute certainty.

51 Holland Michel (n 4), 18.

52 Thomas W Lucas, 'The Stochastic Versus Deterministic Argument for Combat Simulations: Tales of When the
Average Won't Do' (2000) 5 Military Operations Research 9; Timothy J Horrigan, 'Configuration and the Effectiveness of Air Defense Systems in Simplified, Idealized Combat Situations – A Preliminary Examination' (Horrigan Analytics, June 1995), 5.

53 Katrina Wakefield, 'Predictive Modeling Analytics and Machine Learning' (SAS Data and AI Solutions)
<https://www.sas.com/en_gb/insights/articles/analytics/a-guide-to-predictive-analytics-and-machinelearning.html> accessed 10 February 2024.

54 The Pecan Team, 'Contrasting Generative AI, Predictive AI, and Machine Learning' (*Pecan*, 6 December 2023) <https://www.pecan.ai/blog/generative-ai-predictive-ai-machine-learning/>.

55 Avi Goldfarb and Jon R Lindsay, 'Prediction and Judgement: Why Artificial Intelligence Increases the Importance of Humans in War' (2022) 46 International Security 7.

56 Thomas H Cormen et al, *Introduction to Algorithms* (4th edn, MIT Press 2022), 5.

57 Pratap Dangeti, *Statistics for Machine Learning: Build Supervised, Unsupervised, and Reinforcement Learning Models Using Both Python and R* (Packt Publishing 2017), 8.

The accuracy of how often an AI system produces a correct match is referred to as *statistical accuracy* and usually has a small percentage of error. It is reported in a value between 0–1 or 0–100, where 0 demonstrates that it will always predict the wrong label, and 1 and 100 represents that there is always a correct prediction of the correct label. As such, non-deterministic models work with estimates and have a margin of error. The accuracy is used in a *confusion matrix* representing the accuracy of a model and evaluates the model's prediction performance and what errors it is making.[58] The confusion matrix categorises the number of true positives, true negatives, false positives (type I error),[59] and false negatives (type II error).[60] True positives are a complete match and true negatives illustrate that the match is not correct. To illustrate, an algorithm is trained to detect lawful targets: a true positive match represents lawful targets, and unlawful targets represents true negatives. In order to function effectively, an algorithm must have *robustness* to maintain its performance[61] and not be vulnerable against adversarial attacks.[62] Moreover, a system's *reliability* is determined by its trustworthiness and to what extent failures and unintended effects occur. Reliability measures 'how consistently the weapon system will function as intended'.[63]

Increasingly, AI for military use – also referred to as 'war algorithms'[64] – has become an important tool on the battlefield. For example, the US Department of Defense ('DoD') Algorithms-Warfare Team used video feeds from Iraq and Syria captured by drones as input to identify objects and label data.[65] According to the US DoD, traditional collateral damage tools 'cannot always account for the dynamics of the operational environment'[66] in comparison to non-deterministic

58 Aniruddha Bhandari, 'Understanding & Interpreting Confusion Matrix in Machine Learning (Updated 2024)' (*Analytics Vidhya*, 11 January 2024) <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machinelearning/>.

59 ibid. False positives or Type I errors occur when the value was falsely predicted, such that the actual value was negative but the model predicted a positive value.

60 ibid. False negatives or Type II errors occur when the predicted value was falsely predicted, such that the actual value was positive but the model predicted a negative error.

61 Ronan Hamon, Henrik Junklewitz, and Ignacio Sanchez, 'Robustness and Explainability of Artificial Intelligence' (*Publications Office of the European Union*, 2020).

62 To read more about adversarial attacks, see Mark A Visger, 'Garbage in, Garbage Out: Data Poisoning Attacks and Their Legal Implications' in Laura A Dickinson and Edward W Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023).

63 International Committee of the Red Cross, 'Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control' (Geneva, August 2019), 10.

64 Dustin A Lewis, Gabriella Blum, and Naz K Modirzadeh, 'War-Algorithm Accountability' (Harvard Law School Program on International Law and Armed Conflict, 31 August 2016).

65 Lewis, Modirzadeh, and Blum (n 3).

66 US DoD, 'No-Strike and the Collateral Damage Estimation Methodology', (Chairman of the Joint Chiefs of Staff Instruction, CJCSI 3160.01, 12 October 2012).

tools. For example, they are limited in their functions because they do not account for civilian or non-combatant personnel presence in the target area.

3 The Challenges to the Use of FRT and AI in Armed Conflict

This section will analyse the challenges arising from the use of FRT and AI in military targeting situations. It will examine these challenges in the following order: (a) the (in)accuracy of FRT, (b) automation bias, (c) the impact of technical and cognitive biases, and (d) the opacity of AI.

3.1 *The (in)accuracy of FRT*

FRT is an especially difficult biometric because faces are complex and multidimensional. It has been claimed to be one of the least accurate biometric modalities.[67] A persons' face is never static, but its surface changes considerably due to factors such as ageing,[68] makeup,[69] or disability.[70] In comparison, other biometric modalities, such as iris or retina scans, are more accurate because they feature a close contact collection process, whereas face recognition is captured from a distance.[71]

Facial recognition systems can be used in different environments. In *controlled* environments, a face is recognised when factors such as angles and light are more controlled, for example in passport controls.[72] In *uncontrolled* environments, these factors cannot be monitored consistently. Individuals may not be standing still, and in public spaces that are not well-lit. Research has found that the accuracy of FRT in uncontrolled environments is significantly challenged because there is less possibility that the person is looking directly into the camera.[73] Another challenge to the accuracy of FRT is whether footage

67 Mary Clark, 'Top Five Biometrics (Face, Fingerprint, Iris, Palm and Voice) Modalities Comparison' <https://www.bayometric.com/biometrics-face-finger-iris-palm-voice/> accessed 17 February 2024.

68 Leila Boussaad and Aldjia Boucetta, 'Deep-Learning Based Descriptions in Application to Aging Problem in Face Recognition' (2022) 34 Journal of King Saud University – Computer and Information Sciences 2975.

69 Sayako Ueda and Takamasa Koyama, 'Influence of Make-up on Facial Recognition' (2010) 39 Perception 260.

70 European Union Agency for Fundamental Rights (n 36).

71 John D Woodward et al, *Army Biometric Applications: Identifying and Addressing Sociocultural Concerns* (RAND 2001), 19.

72 David Bolt, 'An Inspection of the Policies and Practices of the Home Office's Borders, Immigration and Citizenship Systems Relating to Charging and Fees' (Independent Chief Inspector, 2019).

73 Patrick Grother, Mei Ngan and Kayee Hanaoka, 'Facial Recognition Technology Evaluation (FRTE): Part 2: Identification' (National Institute of Standards and Technology, September 2023) <https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf> [30 January 2025].

is live or non-live. Live Facial Recognition Technology ('LFRT') extracts a face from video footage to identify whether a person exists within a database of images. LFRT is more likely to give false positives[74] because it cannot control factors such as distance, angles, and light.[75]

In the civilian domain, there have been recent cases of false positives when FRT has been used by law enforcement to identify suspects. For example, in a report on the use of LFRT by British law enforcement, it was found that on average the false recognition rate was 95% across the country.[76] The same inaccuracies are reflected in the US. For example, in 2019, the New Jersey police imprisoned a suspect based on an inaccurate facial recognition that had identified him as another man.[77] He was imprisoned for 10 days.[78]

In the context of armed conflict, with civilians present, dust and inadequate lighting conditions make it difficult for FRT systems to accurately identify faces. In particular, when facial recognition systems have not been trained on datasets that represent the diversity of faces in conflict zones, this can lead to higher error rates for underrepresented groups. Further, false positive matches are particularly alarming when they falsely 'match' a civilian as a known person of an armed group. Misidentification in armed conflict can have more serious consequences than in standard law enforcement settings, and it is therefore important to ensure a standard is set for accepting less false positives in armed conflict contexts.

Yet even with a lower false positive identification rate, the disastrous consequences for individuals cannot be ignored. For instance, if FRT is used to identify targets in a group of 200,000 people with a false positive identification rate of 0.1%, then 2000 people would be wrongly identified as a target. Hence,

74 European Union Agency for Fundamental Rights (n 36).

75 Gates (n 32), 71.

76 Big Brother Watch, 'Face Off: The Lawless Growth of Facial Recognition in UK Policing' (May 2018).

77 Kashmir Hill, 'Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match' (*The New York Times,* 29 December 2020) <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>.

78 Many other cases of misidentification have been reported: see Kashmir Hill, 'Wrongfully Accused by an Algorithm' (*The New York Times*, 24 June 2020) <https://www.nytimes.com/2020/06/24/technology/facialrecognition-arrest.html>; Kashmir Hill, 'Eight Months Pregnant and Arrested After False Facial Recognition Match' (*The New York Times*, 6 August 2023) <https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html>; Johana Bhuiyan, 'Facial Recognition Used After Sunglass Hut Robbery Led to Man's Wrongful Jailing, says suit' (*The Guardian*, 23 January 2024) <https://www.theguardian.com/technology/2024/jan/22/sunglass-hut-facialrecognition-wrongful-arrest-lawsuit>.

the accuracy assessment must be determined on error rates, population size, and the sensitivity of the environment.

3.2 *Automation Bias*

Scholars have argued that one of the benefits of AI-based systems is that they can eliminate human errors and are less likely to make mistakes. Therefore, some argue that these machines are to be trusted more than humans given that they do not act from emotional responses and are superior to human capabilities.[79] On the other hand, humans tend to overly trust computer-based decision support systems. They ignore or do not search for contradictory information beyond what the machine informs them.[80] Recent examples illustrate how people overly trust AI: in healthcare doctors have reached inaccurate diagnoses;[81] drivers have crashed vehicles into a destroyed bridge, resulting in fatality;[82] and students have followed instructions by a robot which led them into a burning building.[83] Algorithmic recommendations by AI-DSS can lead to a human response – *automation bias* – where humans favour their automated recommendations while ignoring contradictory information that may suggest the opposite.[84] Research has shown that the risk of automation bias increases in time critical situations.[85]

The second issue is the speed of AI-DSS producing recommendations which are on a far larger scale than human capabilities.[86] There may be a tendency for military commanders to act more quickly and trust recommendations because the constant inflow received in real-time increases a sense of 'urgency'. The

79 Robin Geiß and Henning Lahmann, 'Autonomous Weapons Systems: A Paradigm Shift for the Law of Armed Conflict?' in Jens David Ohlin (ed), *Research Handbook on Remote Warfare* (Edward Elgar Publishing 2017), 373.

80 Elke Schwarz, 'Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control' (2021) 1 The Philosophical Journal of Conflict and Violence 53.

81 Thomas Dratsch et al, 'Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance' (2023) 307 Radiology 1.

82 Jenny Gross, 'He Drove Into a Creek and Died. His Family Blames Google Maps' (*The New York Times*, 21 September 2023) <https://www.nytimes.com/2023/09/21/us/google-maps-lawsuit-collapsed-bridge.html>.

83 John Toon, 'In Emergencies, Should You Trust a Robot?' (*Georgia Tech*, 29 February 2016) <https://news.gatech.edu/news/2016/02/29/emergencies-should-you-trust-robot>.

84 Linda J Skitka, Kathleen Mosier, and Mark D Burdick, 'Accountability and Automation Bias' (2000) 52 International Journal of Human-Computer Studies 701.

85 Mary L Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (2004) American Institute of Aeronautics and Astronautics.

86 Berenice Boutin, 'Legal Questions to the Use of Autonomous Weapons Systems' (Briefing Paper to the AIV/CAVV Advisory Report on Autonomous Weapon Systems: The Importance of Regulation and Investment, 2021), 4.

more complex and automated a system is, the more military commanders may trust their output because they do not feel as if they have time to verify the target. This can affect the extent to which military commanders retain their human judgement when relying on AI-DSS-recommendations in order to take precautionary measures under Article 57 of Additional Protocol ('AP') I.[87] Especially in densely populated areas with a higher civilian presence, it is pivotal that decision-makers take sufficient time to make the necessary assessments to reduce civilian harm.

3.3 *The Impact of Technical and Cognitive Biases*

There are concerns about how AI-based systems may be impacted by biases and how existing biases can be reinforced, developed in the training phase, and occur during the use of the system.[88] Research has found that there are various types of technical biases that are embedded within the data[89] that can affect the interaction with AI-produced recommendations.[90] Firstly, *algorithmic bias* refers to systematic and repeated errors in its outcomes due to unrepresentative data during the training phase. It produces incorrect outcomes because it is more accurate in identifying one particular group than other groups.[91] Secondly, *sampling bias* occurs when some groups of a population are more likely to be selected than others.[92] Thirdly, *group attribution bias* for the 'out-group' happens when stereotyping people who do not belong to a certain group, whereas 'in-group' is when favouring a particular group.[93] Lastly, *action bias* refers to when humans favour action over inaction

87 Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims in International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (Protocol) art 57 ('AP I').

88 ICRC, 'Artificial Intelligence and Machine Learning in Armed Conflict: A Human Centred Approach' (2020) 102 International Review of the Red Cross 463.

89 Lindsey Jacques, 'Facial Recognition Technology and Privacy: Race and Gender – How to Ensure the Right to Privacy is Protected' (2021) 23 San Diego International Law Journal 111; Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) 81 Proceedings of Machine Learning Research 1.

90 Lucía Vicente and Helena Matute, 'Humans Inherit Artificial Intelligence Biases' 13 Scientific Reports 1.

91 Megan Garcia, 'Racist in the Machine: The Disturbing Implications of Algorithmic Bias' (2016) 33 World Policy Journal 111.

92 Andrew D Selbst, 'Disparate Impact in Big Data Policing' (2017) 52 Georgia Law Review 109, 134–135.

93 Nils Karl Reimer et al, 'Self-Categorization and Social Identification: Making Sense of Us and Them' in Derek Chadee (ed), *Theories in Social Psychology* (2nd edn, Wiley-Blackwell 2020); Miles Hewstone, Mark Rubin and Hazel Willis, 'Intergroup Bias' (2002) 53 Annual Review of Psychology 575, 576.

because we can feel compelled to act, even if there is no evidence or we do not have all the necessary information about the output.[94]

Algorithmic bias can be embedded in both FRT and AI-based systems. A study carried out by the Massachusetts Institute of Technology ('MIT'), called the *Gender Shades* project, found that data in the training phase had been 'fed' with pictures of white individuals causing the system to perform worse when identifying people of colour. The researchers found that gender classification algorithms against dark-skinned females had error rates up to 34% higher than lighter-skinned males.[95] Algorithmic bias can impact the level of accuracy of these algorithms when engaging in the 'filtering process'; a process which depends on the quality of the training data. This can occur if an AI-based system is trained on data that is overly focused on one group of people rather than having a more diverse training data. For example, assume a machine learning model is fed with videos and images of people of colour subject to sampling bias, because the developer believes that this group are likely to be terrorists due to racial prejudice.[96] During the training phase, the algorithms will be taught to disproportionately label that group as 'valid' lawful targets far more frequently than other groups. The consequences of these biases in the system increase the likelihood of misidentification and errors to identify civilians as lawful targets.

Recent examples have shown how law enforcement using FRT disproportionality arrests more black people which leads to the wrongful arrest of innocent individuals.[97] In targeting situations, algorithmic bias can thus affect the ability of AI-DSS to distinguish between lawful and unlawful targets.[98] Group attribution bias is particularly problematic in targeting operations. For example, a study from 2009 found that US cadets decided to shoot more rapidly when they were shown images of 'Middle Eastern men wearing traditional clothing'.[99] The soldiers did this because they stereotyped Middle Eastern men as terrorists or enemy combatants.[100] Concluding, there is

94 Michael Bar-Eli et al, 'Action Bias Among Elite Soccer Goalkeepers: The Case of Penalty Kicks' (2007) 28 Journal of Economic Psychology 606.

95 Buolamwini and Gebru (n 89).

96 Kevin K Fleming, Carole L Bandy, and Matthew O Kimble, 'Decisions to Shoot in a Weapon Identification Task: The Influence of Cultural Stereotypes and Perceived Threat on False Positive Errors' (2010) 5 Social Neuroscience 201.

97 Thaddeus L Johnson and Natasha N Johnson, 'Police Facial Recognition Technology Can't Tell Black People Apart' (*Scientific American*, 18 May 2023) <https://www.scientificamerican.com/article/police-facial-recognitiontechnology-cant-tell-black-people-apart/>.

98 Ashley Deeks, 'Predicting Enemies' (2018) 104 Virginia Law Review 1529, 1577.

99 Fleming, Bandy and Kimble (n 96).

100 Keith B Payne and Joshua Correll, 'Race, Weapons, and the Perception of Threat', in Bertram Gawronski (ed), *Advances in Experimental Social Psychology* (Elsevier Academic Press 2020).

a significant likelihood of more targeting errors when pre-existing biases and stereotypes of particular groups are reinforced.[101] More importantly, the impact of action bias may *exacerbate* existing pressure on military commanders to make decisions when relying on AI-recommendations.

3.4 *Opacity of AI*

AI-based systems are generally unknown – *opaque* – for developers and users attempting to understand their process and outcomes. There are various forms of opacity that impact the comprehensibility of AI.[102] The impact of *transparency* refers to the processes in the design, training, testing, and development of the AI-based system. By contrast, *interpretability* refers to a user's ability to understand and predict their performance by being able to explain why certain outputs are generated.[103] Another form of opacity is *traceability*. It refers to the user's or developer's ability to trace 'back' AI-recommendations to investigate their outcomes.[104] These forms of opacity pose challenges in the human-machine interaction, verification of AI-produced recommendations, and questions of responsibility for war crimes.

Governments and private companies are usually not transparent regarding what data they have used to train the algorithms.[105] Moreover, the lack of interpretability raises questions as to whether a military commander can understand the AI-DSS recommendations, specifically why certain targets are labelled as lawful targets and not others. Understanding AI-DSS recommendations requires considerable specialist skill, technical knowledge, and costly resources.[106] Even to their creators, but especially for users, AI-DSS recommendations 'can be black boxes'.[107] Due to a lack of interpretability, AI-DSS users do not find out why certain individuals have been generated as

101 Nema Milaninia, 'Biases in Machine Learning Models and Big Data Analytics: The International Criminal and Humanitarian Law Implications' (2020) 102 International Review of the Red Cross 199.

102 It is important to note that these terms lack a universal definition, and their meanings can vary depending on the author.

103 For more details about interpretability, see Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1, 1.

104 Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829, 1832; Arthur Holland Michel, 'The Black Box, Unlocked: Predictability and Understandability in Military AI' (United Nations Institute for Disarmament Research 2020).

105 Brent Mittelstadt, 'Interpretability and Transparency in Artificial Intelligence' in Carissa Véliz (ed), *The Oxford Handbook of Digital Ethics* (Oxford University Press 2022).

106 Burrell (n 103), 4.

107 Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 Harvard Journal of Law & Technology 890, 891.

recommendations for targeting. Moreover, issues related to the traceability can cause problems during investigations of war crimes.[108] This is known as the 'accountability-gap' due to the opacity, complexity, and unpredictability of AI-based systems.[109]

AI-DSS are opaque as they do not explain why certain individuals are recommended as members of an armed group. Unless DSS-users are offered explanations, there is risk that a DSS-user may make an insufficiently informed decision. It is pivotal for a military commander to know when (or when not) to trust recommendations generated by AI, especially when the consequences of acting upon those recommendations may result in casualties or even fatalities. Additionally, due to the lack of interpretability, it is likely that 'the creator of AI cannot necessarily foresee how the AI will make decisions, what conduct it will engage in, or the nature of the patterns it will find in data, what can be said about the reasonable person in such a situation'.[110] Because algorithmically-generated recommendations are generated on a faster scale than human capabilities, military commanders can make disastrous decisions that potentially lead to violations of IHL.

4 The Law of Targeting

This section will focus on the application of IHL, specifically the law of targeting, to the use of AI-DSS and FRT in military targeting decisions. The law of targeting contains rules concerning under what circumstances persons or objects may be attacked in armed conflicts. This section will contribute to understanding how the use of AI-DSS and FRT may or may not be compliant with the rules of targeting, while using the Lavender as a case study. It will examine (i) the principle of distinction, (ii) the principle of proportionality, and (iii) the principle of precautions in attack.

4.1 *The Principle of Distinction*

The principle of distinction has been recognised as a 'cardinal principle' of IHL by the International Court of Justice ('ICJ').[111] It is considered to constitute

108 Marta Bo, Laura Bruun, and Vincent Boulanin, 'Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS' (SIPRI, October 2022).

109 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) 19 Journal of International Criminal Justice 275.

110 Bathaee (n 107), 924.

111 *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) 1996 ICJ Rep. 226, 257.

a rule of customary international law ('CIL').[112] The rule has been codified in Article 48 of AP I:

> In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.

The rule requires that parties to an armed conflict must distinguish between lawful and unlawful targets, and to only direct attacks against the former. It is prohibited to directly target civilians or the civilian population,[113] unless a civilian loses their protected status by directly participating in hostilities and becoming classified as a civilian DPH.[114] Civilian objects shall not be the object of an attack and are defined in negative terms, as objects that do *not* constitute military objects.[115] Military objects are defined as 'those objects which by their nature, location, purpose, or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage'.[116] The definition sets out two cumulative conditions that complement each other.

Using FRT could provide a more effective and accurate identification of individuals, depending on the level of accuracy of such enrolled biometrics. If the Lavender receives input from FRT, it is able to search, detect, and identify pre-enrolled individuals in a biometric database. Scholars have argued that using FRT to identify individuals could facilitate adherence to the principle of distinction.[117] Moreover, IHL does not seem to directly prohibit the use of FRT.[118] Yet, there could be rules of IHL that may restrict or prohibit certain categories of persons to be identified by FRT.[119]

As previously mentioned, the Israeli military has relied on facial recognition to identify individuals in Gaza. In the context of the conflict in Gaza, FRT may be used to *identify* whether a particular individual is a previously known fighter

112 Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Volume 1: Rules* (CUP 2005) 62, rule 7.
113 AP I (n 87), art 51; Henckaerts and Doswald-Beck (n 112), rule 1.
114 AP I (n 87), art 51(3); Henckaerts and Doswald-Beck (n 112), rule 6.
115 AP I (n 87), art 52(1); Henckaerts and Doswald-Beck (n 112), rule 8.
116 AP I (n 87), art 52.
117 Boothby (n 8), 397; Zwanenburg (n 43), 1416; Mitchell (n 7), 305–306.
118 Mitchell (n 7), 306.
119 Zwanenburg (n 43); Emily Crawford, 'The Right to Privacy and the Protection of Data for Prisoners of War in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE Publications 2022).

for Hamas. FRT has relevant use for identifying or verifying the identity if an individual who has previously been enrolled in a biometric system, but whose status under IHL cannot be determined. For instance, AI-DSS in combination with FRT can be used in addition to intelligence information that informs whether an individual is a known Hamas fighter. As indicated earlier, ensuring accuracy when identifying an individual by FRT may be especially cumbersome in complex and uncontrolled environments. In densely populated areas, like in Gaza, it is likely that there are difficulties for a facial recognition system to identify individuals, because it cannot properly detect faces due to difficult angles and shadows. Therefore, false positives are likely, wherein the system may incorrectly identify an individual as 'matching' a person enrolled in the system as a known fighter for Hamas. According to one military IDF official, '[a]t times, the technology wrongly flagged civilians as wanted Hamas militants'.[120] Therefore, the use of AI-DSS with input from FRT to identify individuals in densely populated areas may be less effective and lead to an increase of false positive identifications which can cause harm to the civilian population.

The principle of distinction requires a contextual interpretation of who is and who is not a civilian that may not be easily translated into a machine. The assessment requires a complex analysis of what movements and actions would belong to a civilian or a combatant.[121] As Christof Heyns explains, a system must differentiate between a 'civilian with a large piece of metal in his hands' and 'a combatant in plain clothes'.[122] In IHL, it is lawful to target civilians who are DPH because they have lost their protection from direct attacks whilst retaining civilian status.[123] IHL does not define what DPH constitutes, but sets out that civilians are immune from attacks 'unless and for such time as they take a direct part in hostilities'.[124] This has been accepted as CIL.[125] Nevertheless, what exactly constitutes DPH remains a point of controversy among scholars.[126] The Interpretative Guidance on the Notion

120 Frenkel (n 16).

121 For more extensive details, see ICRC, 'ICRC Position on Autonomous Weapon Systems' (Geneva, 12 May 2021), 9.

122 Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (Human Rights Council) A/HRC/23/47 (9 April 2013), 67.

123 ICRC, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 in Yves Sandoz, Christopher Swinarski and Bruno Zimmermann (eds) (1987) ('ICRC Commentary'), para 1942.

124 AP I (n 87), art 51(3).

125 Henckaerts and Doswald-Beck (n 112), rule 6; The Supreme Court of Israel, *The Public Committee against Torture in Israel et al v. The Government of Israel et al* (Judgment) Case No. HCJ 769/02 (11 December 2006), 50.

126 Marco Sassòli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90

of DPH by the International Committee to the Red Cross ('ICRC') has set out an interpretative guide to clarify when a civilian is DPH.[127] An assessment of the three constitutive elements of whether a specific act[128] constitutes DPH in 'the circumstances prevailing at the relevant time and place' requires a context-specific analysis.[129] This assessment is not a numerical one, but rather requires a qualitative analysis to determine whether an individual is DPH by considering their intentions. It is difficult to comprehend how this analysis could be effectively translated into machine coding to categorise, detect patterns, and combine all relevant elements to accurately distinguish civilians from individuals who DPH.[130]

In targeting situations, militaries use AI-DSS that gather data and filter information to differentiate between lawful and unlawful targets. This analysis is based on pattern-matching. As such, some AI-DSS rely on so-called *assumptions* that transform data into information (the output).[131] However, while AI-DSS can be designed to identify correlations, they cannot establish causation, as this is an impossible mathematical analysis to perform.[132] For example, the Lavender has allegedly been designed to learn how to 'identify characteristics of known Hamas and PIJ operatives.'[133] The machine learning model was fed with information and data to assess and provide a ranking from 1 to 100 of the likelihood that a person is a member of Hamas or PIJ. One of the IDF officers explained that the Lavender has 'sometimes mistakenly flagged individuals who had communication patterns similar to known Hamas or PIJ operatives'.[134] While it is difficult to verify this information due to lack

International Law Studies 308; Michael N Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to Critics' (2013) Harvard National Security Journal Feature; Kenneth Watkin, 'Opportunity Lost: Organized Armed Groups and the ICRC 'Direct Participation in Hostilities' Interpretive Guidance' (2010) 42 International Law and Politics 641.

127 ICRC, Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law (Geneva, 2009).

128 For more details see ibid, 47–50, 51–58, 58–64.

129 ibid, 42.

130 Jeroen van den Boogaard, 'Proportionality and Autonomous Weapons Systems' (2015) 6 Journal of International Humanitarian Legal Studies 247, 262–263; Sassòli (n 126), 328–330.

131 Holland Michel (n 4), 36–37.

132 Hengameh Irandoust and Abder Benaskeur, 'Human-Autonomy Teaming for Critical Command and Control
Functions' (IEEE International Conference on Human-Machine Systems, Rome, 2020), 3; Matteo Pasquinelli and Vladan Joler, 'The Nooscope Manifested: AI as Instrument of Knowledge Extractivism' (2021) 36 AI & Society 1263, 1276.

133 Abraham (n 12).

134 ibid.

of transparency from Israel, it is pivotal that AI-DSS users are aware of these inherent assumptions to avoid harm to civilians and use appropriate data to distinguish between lawful and unlawful targets.

Some militaries work within the 'OODA loop' (Observe, Orient, Decide, Act). By relying on AI-DSS, it can create a faster OODA loop and accelerate decision-making.[135] Reportedly, the Lavender is used to analyse information for potential members of Hamas or other armed groups in the target development phase. While the IDF has claimed that the Lavender is not used for identifying or predicting whether persons are terrorists,[136] reports indicate that the Lavender aggregates a variety of sources about individuals, and provides output regarding who may be a member of Hamas. What is particularly concerning is the use of 'Where's Daddy?' which tracks suspected militants and sends a signal to the IDF when they have entered their home. At times, the IDF have 'bombed [targets] in homes without hesitation, as a first option'[137], and in certain instances, while their family was still present. Allegedly, both the Lavender and Where's Daddy were used in the targeting process. While civilians who are DPH are lawful targets under IHL, to bomb an entire family without first verifying whether each member is classified as a civilian DPH is seriously concerning and unlikely to be consistent with the principles of distinction, proportionality, and precautions in attack. As noted above, the output produced by the Lavender is reviewed firstly by intelligence analysts.[138] In 2023, an IDF official explained that human analysts no longer need to review one single target for hours as it 'now takes minutes, with a few more minutes for human review'.[139] If approved, reviewed targets are transferred to those responsible for planning and executing attacks. According to Mimran and Gal, these targets are sent to 'a target room' where legal advisors, operational advisors, and senior intelligence officers revise targets based on IHL principles.[140] While the AI-DSS output does not provide a classification of an individual's status under IHL, it can inform the user's IHL categorisation. Yet, FRT can be used to identify or verify an individual's identity based on pre-enrolled biometrics, but is unable to determine an individual's status under IHL. Therefore, this paper argues that users relying on both AI-DSS recommendations and FRT must perform

135 Owen Daniels, 'Speeding Up the OODA Loop with AI: A Helpful or Limiting Framework?' 2021 Joint Air & Space Power Conference, 159.
136 IDF Website (n 14).
137 Abraham (n 12).
138 ibid.
139 Newman (n 21).
140 Mimran and Dahan (n 22).

a legal assessment to determine an individual's status under IHL. Given the potential for certain sources of information to be overlooked by the AI-DSS, it is pivotal that personnel ensure that unlawful targets are distinguishable from lawful targets in the target selection phase. Because the output of the Lavender depends upon the accuracy and reliability of the algorithms, it is crucial that there are sufficient personnel to undertake an analysis to review the accuracy of these recommended targets and outputs from FRT, receive additional information if necessary, and ensure compliance with the principle of distinction.

4.2 *The Principle of Proportionality*

The principle of proportionality in IHL serves as a safeguard both against indiscriminate attack,[141] and as a precautionary measure requiring that those who plan or decide upon attacks must take to refrain from such attacks that violate the proportionality rule.[142] The rule is considered CIL[143] and the principle is codified in Article 51(5)(b) of AP I:

> [A]n attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.

The principle requires that those who plan and carry out operations must consider the potential effects of that operation before the attack, such as death or injury to civilians and destruction of civilian property. As such, the principle adds an additional restraint to that imposed by the principle of distinction. The rule sets out the obligation that those responsible must assess those incidental effects of planned attacks – known as 'collateral damage' – and identify those which are excessive to the direct military advantage being sought by the operation. Civilians and civilian objects must be spared from collateral damage to the greatest possible extent. The principle recognises that there is a risk of incidental loss of life or injury in war, and a military attack may be lawful as long as expected collateral damage is not excessive to the concrete and direct military advantage anticipated.[144] The rule requires weighing (a) expected incidental loss compared to the (b) anticipated military advantage.

141 AP I (n 87), art 51(5)(b).

142 AP I (n 87), art 57(2)(a)(iii).

143 Henckaerts and Doswald-Beck (n 112), rule 14.

144 Amichai Cohen and David Zlotogorski, *Proportionality in International Humanitarian Law: Consequences, Precautions, and Procedures* (Oxford University Press 2021), 4.

Further, the concrete and direct military advantage 'must be perceived in a contextual fashion'[145] and the concept of military advantage must be concrete and not hypothetical.[146] There are diverse views regarding what constitutes a military advantage.[147] The more accepted interpretation is 'in view of the *attack as a whole*', rather than isolated or specific attacks.[148] Scholars stress that 'what is meant by "excessive" is an extremely elusive concept to define and to apply'.[149] The other pivotal concept in the proportionality assessment is 'incidental harm' to civilians.[150] Incidental harm concerns loss of life, mental harm, destruction of civilian objects, damage to the environment,[151] and reverberating effects.[152] Moreover, Article 57(2)(a)(iii) of AP I sets out that the obligation to comply with this rule lies with those who plan or decide upon the attack. While the 'reasonable commander standard' is not defined by IHL, it has been described as 'the standard against which a decision on proportionality is to be made or judged'.[153]

As previously discussed, the +972 Magazine and Local Call report that the Lavender is used in conjunction with the Gospel. It is designed to calculate expected casualties in each attack.[154] The number of casualties – collateral damage estimates ('CDE') – are calculated before attacks and the IDF units are made aware of the expected number of casualties prior to each attack. AI-DSS providing CDE can enhance accuracy by balancing the need for military advantage whilst also minimising incidental loss of life.[155]

145 Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (3rd edn, Cambridge University Press 2016), 108.

146 Nils Melzer, *Targeted Killing in International Law* (Oxford University Press 2009), 293.

147 For a narrower interpretation see, ICRC Commentary (n 123), para 2218. For broader interpretations see Judith Gardam, *Necessity, Proportionality and the Use of Force by States* (Cambridge University Press 2004), 101; Program on Humanitarian Policy and Conflict Research (HPCR), *Commentary to the HPCR Manual on International Law Applicable to Air and Missile Warfare* (Cambridge University Press 2013), 45.

148 Cohen and Zlotogorski (n 144), 66.

149 Michael Wells-Greco, 'Operation 'Cast Lead': *Jus in Bello* Proportionality' (2010) 57 Netherlands International Law Review 397, 399.

150 ICRC Commentary (n 123), para 1913; Geoffrey Corn and Andrew Culliver, 'Wounded Combatants, Military Medical Personnel, and the Dilemma of Collateral Risk' (2017) 45 Georgia Journal of International and Comparative Law 445.

151 Cohen and Zlotogorski (n 144), 78–82.

152 For more details, see Ian Henderson and Kate Reece, 'Proportionality under International Humanitarian Law: The 'Reasonable Military Commander' Standard and Reverberating Effects' (2018) 51 Vanderbilt Journal of Transnational Law 835.

153 ibid, 840.

154 Abraham (n 12).

155 Sassòli (n 126).

Yet, it is a serious concern that the IDF has reportedly been 'loosening constraints regarding expected civilian casualties'.[156] On the one hand, the IDF is emphasising the enhanced precision in targeting, that is enabled by the rapid and automatic extraction of intelligence to generate targets.[157] The IDF has made a statement that once a target has been approved for attack, they conduct an individual assessment per strike.[158] On the other hand, one of the interviewed IDF officials in the targeting operation room stated that '[i]n practice, the principle of proportionality did not exist'.[159] Another IDF official explained that, in some attacks, the IDF authorised the killing of hundreds of civilians in the pursuit of targeting senior ranking Hamas commanders.[160] More generally, the use of AI-DSS can improve adherence to the principle of proportionality by providing more information about the potential number of casualties in attacks. Yet, there are significant concerns as to whether the IDF is effectively using available information to assess expected collateral damage and ensure that it remains proportionate to the anticipated military advantage, consistent with the principle of proportionality.

However, scholars disagree as to how to define the concept of proportionality and to what extent AI-DSS can be useful in proportionality assessments. Some argue that it remains unclear whether and how these complex, context-based, and value-based requirements can be operationalized into a mathematical formula.[161] Others argue that considering the fast technological developments in this area, it may be possible to pre-programme an implementation of the

156 Abraham (n 23).

157 ל הפועל מסביב לשעון'צה של המטרות למפעל הצצה [A Glimpse of the IDF's Target Factory that Operates Around the Clock] (*Israel Defense Forces*, 2 November 2023) <https://www.idf.il/%D7%90%D7%AA%D7%A8%D7%99%D7%99%D7%97%D7%99%D7%93%D7%95%D7%AA/%D7%99%D7%95%D7%9E%D7%9F-%D7%94%D7%9E%D7%9C%D7%97%D7%9E%D7%94/%D7%9B%D7%9C-%D7%94%D7%9B%D7%AA%D7%91%D7%95%D7%AA/%D7%94%D7%A4%D7%A6%D7%95%D7%AA/%D7%9E%D7%9C%D7%97%D7%9E%D7%94-%D7%9E%D7%98%D7%A8%D7%95%D7%AA-%D7%A9%D7%94%D7%95%D7%AA%D7%A7%D7%A4%D7%95-%D7%9B%D7%95%D7%97%D7%95%D7%AA-%D7%A6%D7%94-%D7%9C-%D7%90%D7%92%D7%A3-%D7%94%D7%9E%D7%95%D7%93%D7%99%D7%A2%D7%99%D7%9F-%D7%97%D7%99%D7%9C-%D7%94%D7%90%D7%95%D7%95%D7%99%D7%A8-%D7%97%D7%99%D7%9C-%D7%94%D7%99%D7%9D/>.

158 IDF Website (n 14).

159 Abraham (n 12).

160 ibid.

161 Robert D Sloane, 'Puzzles of Proportion and the Reasonable Military Commander: Reflections on the Law, Ethics, and Geopolitics of Proportionality' (2015) 6 Harvard National Security Journal 299, 322–323; Cohen and Zlotogorski (n 144), 59.

principle of proportionality into machine coding.[162] Additionally, some argue that IHL 'does not require subjective value judgments that machines are unable to make, but depends on an objective assessment of facts'.[163] Others raise the question whether algorithms for use in war are able to balance incidental harm against anticipated military advantage.[164] Moreover, military advantage and incidental harm to civilians 'cannot be compared through the simple use of a formula, as there is no common denominator between them'.[165] As a whole, there are difficulties in 'quantifying the factors of the equation'[166] as the process 'is a singularly subjective and indeterminate legal standard'.[167] As Yoram Dinstein observes:

> Military advantage and civilian casualties/damage are incomparable in a quantifiable manner, and they cannot be configured in a manner resulting in an arithmetical common denominator. Projected civilian losses may be calculated, just as civilian damage may be estimated; but how can one appraise an anticipated military advantage on a measurable scale? The incommensurability of military advantage and civilian casualties/damage often vitiate an objective balancing act between the two.[168]

I argue that proportionality assessments are subjective and subject to a broad range of judgement.[169] The assessment is based on the weighing of two different values: (i) anticipated military advantage, and (ii) expected civilian casualties or damage. While the number of civilian casualties is likely to be quantified, that is not the whole test of proportionality. The text of AP I sets out that one must assess whether civilian damage and injury would be 'excessive *in relation to*' the concrete and direct military advantage anticipated. Therefore, a predetermined amount of civilian casualties or civilian damage cannot be used to accurately evaluate the principle of proportionality. An attack may

162 Schmitt (n 126); Sassòli (n 126).
163 Sassòli (n 126), 339.
164 van den Boogaard (n 130), 267.
165 Cohen and Zlotogorski (n 144), 59.
166 Michael Bothe, Karl Joseph Partsch and Waldemar Solf, *New Rules for Victims of Armed Conflicts Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (2nd edn, Martinus Nijhoff Publishers 2013), 227.
167 Sloane (n 161), 301–302.
168 Dinstein (n 145), 158.
169 Yuval Shany, 'Toward a General Margin of Appreciation Doctrine in International Law?' (2005) 16 European Journal of International Law 907; Luke Whittemore, 'Proportionality Decision Making in Targeting: Heuristics, Cognitive Biases, and the Law' (2016) 7 Harvard National Security Journal 577.

not violate the principle unless the collateral damage is excessive in relation to the concrete and direct military advantage. The balancing between these two concepts is a subjective judgement, rather than an objective one.[170] It is unclear if and how these concepts can be converted and operationalized in machine coding, due to the complexity of these concepts and because they are not clearly defined amongst scholars. Moreover, the ICRC Commentary indicates that while it is 'based to some extent on a subjective evaluation, the interpretation [of the principle of proportionality] must above all be a question of common sense and good faith for military commanders.'[171] Ultimately, the onus rests upon the military commander, not an AI-DSS, to assess the effects of proposed attacks and 'carefully weigh up the humanitarian and military interests at stake'[172] because it is the military commander that is personally responsible for adhering to the principle of proportionality. Therefore, the military commander must be aware of the possibility of failure regarding outputs from a system. Military commanders should not 'be under the impression that these values in any way constitute ground truth, an exact science, or flawless data'.[173] Assume, for example, that the Gospel provides a wrongful estimation of collateral damage that in reality is higher than estimated. It is the responsibility of the military commander to maintain the capacity for making correct assessments regardless of whether the information that the Gospel is produces is correct.

This raises a question as to how the 'reasonable military commander' standard can be applied to the interplay between algorithmically-generated recommendations and the individual military commander. In the Final Report of the International Criminal Tribunal committee reviewing a Bombing in Yugoslavia, published in 2000, it was acknowledged that 'the determination of relative values must be that of the "reasonable military commander"'.[174] Further, in the *Galić* case, the International Criminal Tribunal for the Former Yugoslavia ('ICTY') said that:

> In determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances

170 Cohen and Zlotogorski (n 144), 59.
171 ICRC Commentary (n 123), para 2208.
172 ibid.
173 US DoD (n 66).
174 Office of the Prosecutor, 'International Criminal Tribunal for the Former Yugoslavia: Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia' ('Final Report') (2000) 39 International Legal Materials 1257, 50.

> of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack.[175]

As has been explored above, a military commander must not 'turn a blind eye on the facts of the situation; on the contrary, he is obliged to take into account all available information'.[176] Moreover, what might be reasonable to one commander might be unreasonable to another. Contrary views in the assessment of what is 'reasonable' may complicate issues surrounding the assignment of criminal liability to military commanders. For example, a 2020 study asked a group of legal experts from 11 countries, two military officers, and various laypeople to evaluate proportionality based on several factors. The study found that the academic and military experts were not able to reach agreement as to what constituted the maximum acceptable loss of civilian life.[177] Further, previous research has suggested that individual human factors may influence the 'reasonable military commander', such as the 'background and values of the decision maker'[178] and 'different doctrinal backgrounds and differing degrees of combat experience or national military histories'.[179] However, I argue that there are numerous external factors concerning the 'reasonable military commander' standard that may shape, influence, and impact a military commander's decision making. The first relevant factor is the design and use of the algorithm's recommendations that identify and recommend lawful targets. It is the algorithmically-generated recommendations that provide information that shapes commanders' situational awareness about the battlefield. These recommendations are produced within seconds in real-time. The large volumes of target recommendations produced each day increase the pressure on commanders to act and raises the potential for hasty decisions because it allows little time for individual commanders to make a proportionality assessment. The second relevant factor is the opacity of AI. If neither the intelligence analysts nor military commanders can understand why the AI-DSS classified a person as a lawful target to verify the target's nature, it is difficult for them to make a sufficiently informed decision. Thirdly, there may be limited technical

175 *Prosecutor v Galic* (Trial Chamber) IT-98-29 (5 December 2003), [58].

176 Frits Kalshoven, 'Implementing Limitations on the Use of Force: The Doctrine of Proportionality and Necessity', (1992) 86 Proceedings of the Annual Meeting (American Society of International Law) 39, 44.

177 Daniel Statman et al, 'Unreliable Protection: An Experimental Study of Experts' *In Bello* Proportionality Decisions' (2020) 31 European Journal of International Law 429.

178 Final Report (n 174), 50.

179 ibid.

knowledge and skills to interpret and use AI to inform commanders' decision-making. Does the 'reasonably well-informed' commander making 'reasonable use of information available' take into account the reliance of algorithmically-generated recommendations, opacity of AI and technical knowledge?

I argue that the answer to this question depends upon whether a military commander is able to make reasonable use of the information available from the AI-DSS recommendations. The level of technical understanding required by commanders may vary depending on their role, operational level, and supporting intelligence professionals. Yet, IHL does not regulate the level of technical knowledge of military commanders: what technical training they must take, subsequent level of technical knowledge they must obtain, and ensuring awareness of error rates and causes of failure.[180] However, developments of military AI are likely to become even more complex and intelligent, increasing the opaqueness of these systems.[181] It has been stated that '[i]t may be that certain technology may never meet the ideal levels of transparency desired by regulators and governments'.[182] Consequently, one must further examine what the 'reasonable military commander' standard means in relation to the question: to what extent can military commanders rely on AI-DSS recommendations in targeting decisions?

4.3 *The Principle of Precautions in Attack*

The principle of precautions in attack is codified in Article 57 of AP I and accepted as CIL.[183] The principle requires that in the conduct of military operations, precautionary measures must be taken to protect civilians and civilian objects. The term 'military operations' encompasses 'any movements, manoeuvres and other activities whatsoever carried out by the armed forces with a view to combat' or 'related to hostilities'.[184] Article 57(1) of AP I also requires that parties to the armed conflict take, at all times, 'constant care' to protect the civilian population. This obligation is a continuous one and has no temporal limitations.[185] It has been argued that the duty of constant care

180 For further discussion, see Jonathan Kwik, 'Lawfully Using Autonomous Weapon Technologies: A Theoretical and Operational Perspective' (PhD Thesis, University of Amsterdam [2024]), ch 5.

181 Bathaee (n 107), 929.

182 ibid.

183 Henckaerts and Doswald-Beck (n 112), rule 15.

184 ICRC Commentary (n 123), para 2191.

185 Asaf Lubin, 'The Duty of Constant Care and Data Protection in War' in Laura A Dickinson and Edward Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023).

is a 'general, broad, and flexible duty'[186] and is not limited to certain activities provided under Article 57 of AP I.[187] Therefore, some scholars have suggested that this provision is in itself an obligation due to a broader scope of application for protection,[188] and that the principle creates concrete legal obligations.[189] A close examination of the wording of Article 57(1) of AP I indicates that the scope of the obligation may be broader than those that follow. The first paragraph applies more broadly to 'military operations', which include 'any movements, manoeuvres, and other activities whatsoever carried out by the armed forces with a view to combat'.[190] By contrast, the following paragraphs refers to 'attack', which suggests a narrower scope of application. I agree that the nature of the duty of constant care appears to offer a broader protective scope, as the term 'military operations' is intended to encompass more than just 'attack', including because it suggests imposing legal obligations that may not be related to attacks.

The term 'constant care' has not been defined by IHL. The *Tallinn Manual 2.0* states that the duty of constant care 'requires commanders and all others involved in the operations to be continuously sensitive to the effects of their activities on the civilian population and civilian objects, and to seek to avoid any unnecessary effects thereon'.[191] An analogous interpretation of the obligation can be drawn for AI-DSS. Inherent technical biases in AI-based systems from the development phase could lead to physical harm. For example, if an AI-based targeting system has been trained with predominantly white faces, it is likely to have higher error rates because it is not able to accurately identify people of colour, resulting in physical harm for those people. Previously mentioned research that examined the performance of a facial recognition system found that the software performed far better in identifying white people and had true positive rates up to 99% for white males. The true positive rates decreased to 34% for darker skinned females, because the algorithms had been trained with mostly white male faces.[192] Therefore, I argue that it is important that AI-DSS receiving input from FRT to be used in armed conflicts must be based on data that is representative of the demographic of individuals being targeted.

186 ibid, 236.

187 Eric Talbot Jensen, 'Cyber Attacks: Proportionality and Precautions in Attack' (2013) 89 International Law Studies 198, 202.

188 ibid; Jean-François Quéguiner, 'Precautions under the Law Governing the Conduct of Hostilities' (2006) 88 International Review of the Red Cross 793.

189 Quéguiner (n 189).

190 ICRC Commentary (n 123), para 2191.

191 Michael N Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press 2017), 477.

192 Buolamwini and Gebru (n 89).

Currently, there is a lack of transparency surrounding the development of military AI, and it is not known whether or to what extent racial bias exists in algorithms.[193] The 'constant care' duty may require that States take efforts to ensure that pre-selected targets and biometric data are accurate before, during, and continuously in military operations.

Additionally, the obligation may require that developers of military AI make efforts to acquire representative training data in order minimise the impact of bias and unpredictable outputs to protect the civilian population from harm, injury, and loss of life. The specific requirement to verify targets, as set out in Article 75(2)(a)(i) of AP I, may require a continuous monitoring of data to ensure accuracy. In order to ensure that targeting systems do not malfunction, it is important that they are trained on representative data. However, the existing binding nature of the duty of constant care needs further clarification, considering it remains unsettled how the duty applies to the use of AI-DSS in armed conflicts. Reported by the +972 Magazine and Local Call, the machine learning algorithms were fed with training data that included information about non-military employees within the Hamas government. This included police, civil defence workers, militants' relatives and even Gazan residents sharing the same name identical to a known Hamas militant. This resulted in the Lavender inaccurately 'flagging' civilians as Hamas operatives, because they had similar communication or behavioural patterns as known Hamas operatives.[194] Additionally, it was reported in 2021 that one of the challenges the IDF had 'in deploying this system thus far is that it lacks data with which to train its algorithms on what is not a target'.[195] This is concerning because it reveals that the algorithms have not been developed using *representative* data, but rather they have only 'sampled' data from what constitutes lawful targets. The consequence of sampling bias is that the machine learning model may introduce systematic errors and incorrect recommendations by failing to distinguish between protected and non-protected persons.[196]

As remarked, all stochastic models have a certain margin of error. Reportedly, the Lavender has a 10% margin of error. Therefore, 10% of the 37,000 individuals that were marked as military targets were, in reality, not.[197] While these numbers are difficult to verify, this would mean that 3,700

193 ibid.
194 Abraham (n 12).
195 Gaza Conflict Task Force, 'Gaza Conflict 2021 Assessment: Observations and Lessons' (October 2021) <https://jinsa.org/wp-content/uploads/2021/10/Gaza-Assessment.v8-1.pdf> accessed 1 February 2024, 31.
196 Selbst (n 92), 134–135.
197 Abraham (n 12).

individuals were unlawfully targeted (based on the assumption that all targets were engaged). The IDF, during the conflict, had adjusted 'the bar of what a Hamas operative is'[198] and broadened the scope of a 'Hamas operative'. There is, therefore, concern as to what data has been used to label 'Hamas operatives' in the training phase. The IDF has explained that the 'dataset is regularly updated and its data verified'.[199] Developers update, re-label, and verify the data, to improve the algorithm's performance in the course of the conflict. Unless current users of AI-DSS are informed about how these algorithms have been trained and what labels are used, there may be an exacerbation of AI opacity for military commanders, thereby affecting their ability to understand why certain individuals are recommended as 'Hamas operatives' and make their own decision based on this recommendation. This illustrates the inherent danger to the civilian population of having a high margin of error in AI-DSS, and the role that developers' who design, train, and make choices in the development of these algorithms that are unknown, invisible and not communicated to end-users of AI-DSS in armed conflict. This example illustrates the importance of minimising the impact of AI opacity and carefully updating labels during armed conflicts. To fully grasp the implications of the constant care duty, it is essential to further explore its scope and applicability to the use of AI-DSS.

Article 57(2)(a)(i) of AP I requires that planners and decision-makers do everything feasible to *verify* that the target is a military objective and not subject to special protection. The Lavender is operating at an unprecedented pace and is able to process information rapidly from vast volumes of information. The impact of automation bias in such situations could be that it becomes more difficult for users to cancel attacks considering the rapid pace with which technology generates recommendations. Moreover, it has been suggested that the AI-DSS 'processes a lot of data better and faster than any human, and translates it into targets for attack'.[200] It is therefore likely that human analysts and military commanders may trust these calculations more than themselves and fail to search for contradictory information. Because of this, I maintain that it is important that AI-DSS users take sufficient time to review algorithmically-generated recommendations and slow down the decision-making process to ensure that non-military targets are protected from direct attack. However, it remains unclear how much time must be dedicated to review the accuracy

198 ibid.
199 IDF Website (n 14).
200 Abraham (n 23).

of AI-DSS recommendations to ensure compliance with the precautionary principle.

Moreover, another concern is how and to what extent the dedicated time to review targets may impact compliance with the principle of precautions in attack. In the early stages of the war in October 2023, an intelligence officer explained that they would devote 20 seconds 'to each target before authorizing a bomb'[201] and do dozens of approvals every day.[202] Reportedly, the review was sometimes limited to checking whether the target was a male. The IDF has not disclosed how long they dedicate to conducting a legal review of targets before engagement. If the described review process was implemented in practice, even occasionally, it raises serious IHL concerns. Firstly, that the review process did not involve a thorough assessment of the target's legitimacy under IHL, but rather an analysis of an individual's gender. Secondly, what form of analysis the intelligence analysts conduct and their responsibility at the review-stage.[203] Thirdly, the danger of not taking sufficient time to verify targets. If relying on AI-DSS recommendations, there should be a thorough assessment in terms of the accuracy of an AI-generated target recommendation. Users should be aware of the information and sources that have been relied upon to generate these targets in order to gather, if necessary, other relevant information. Finally, there must be a legal analysis as to whether an individual has, for instance, DPH status and could qualify as a lawful target. Military commanders have the primary responsibility for undertaking precautionary measures as AI-DSS are not intended to replace the decision-maker, but rather *support* decision-making.[204]

The use of FRT can enhance verification of an individual's identity. It can even be argued that those who plan and execute attacks are required to use FRT, if they possess this technology in order to 'do everything feasible to verify' their targets' identity.[205] Therefore, using FRT in targeting decisions might not be controversial, but rather required.[206] Yet, the verification of an individual's identity may provide information regarding whether a person is a civilian or part of an armed group. It can be useful when searching for a specific individual that is known to be part of Hamas. As already established, FRT cannot be used to determine whether a person is a lawful target because it does not identify a

201 ibid.
202 Abraham (n 12).
203 Asaf Lubin, 'The Reasonable Intelligence Agency' (2022) 47 Yale Journal of International Law 120.
204 Sassòli (n 126), 335–336.
205 Boothby (n 8), 400.
206 Boothby (n 8).

person's *status*, only their *identity*. As such, the usefulness of AI-DSS receiving input from FRT seems to be dependent upon the context. It can be more useful in targeting pre-known individuals rather than in major combat operations and densely populated areas. Nevertheless, it is important to take sufficient time when reviewing recommendations, because it may lead to an action bias, given that the Lavender is identifying targets at a far faster pace than human capabilities. With that increased number of targets, it is likely that action bias is exacerbated by the IDF's strategy as reportedly the 'emphasis is on quantity and not on quality'.[207] Moreover, the loosening restraints of collateral damage suggests that action rather than inaction is aligned with the IDF's overall strategy.[208] If relying on AI-DSS recommendations in targeting decisions, it may more 'typically privilege action over non-action in a time-sensitive human-machine configuration'.[209] These recommendations are first reviewed by human intelligence analysts that determine whether to authorise the recommendation for further review. It is conceivable that human analysts are overwhelmed by the high volume of data as well as being in an armed conflict. Another challenge is group attribution bias because AI-DSS produce an overflow of recommendations and human analysts may choose targets that confirm their own biases. The impact of opacity limits to what extent a military commander can verify the accuracy of these recommendations because they cannot understand the process and may not have a sufficient level of confidence in the system. Due to the unprecedented pace and expansion of targets, it raises the concern as to whether decision-makers are able to retain the responsibility to verify the accuracy of provided targets. The quest for speed and quantity offers a decision-advantage at a potential cost of insufficiently reviewing targets. Allowing more time to review algorithmically-generated recommendations could enhance compliance with precautionary measures, given the high complexity of AI, the fast pace of algorithmic recommendations, and the stressful environment in which decision-making occurs.

207 Abraham (n 23).

208 A source explained to the +972 Magazine and Local Call report: 'When a 3-year-old girl is killed in a home in Gaza, it's because someone in the army decided it wasn't a big deal for her to be killed – that it was a price worth paying in order to hit [another] target. We are not Hamas. These are not random rockets. Everything is intentional. We know exactly how much collateral damage there is in every home': ibid.

209 Neil Renic and Elke Schwarz, 'Inhuman-in-the-loop: AI-Targeting and the Erosion of Moral Restraint' (*Opinio Juris*, 19 December 2023) <https://opiniojuris.org/2023/12/19/inhuman-in-the-loop-ai-targeting-and-theerosion-of-moral-restraint/>.

5 Conclusion

This article offers a preliminary discussion to outline the challenges that may arise when military decision-makers use AI-DSS in targeting decisions. Technology is becoming increasingly sophisticated and it is expected that more militaries will develop and employ technologies that promote faster and more efficient decisions.[210] AI-DSS can analyse large amounts of data and generate recommendations to support decision-making.[211] While AI-DSS are not weaponised and do not autonomously 'pull the trigger', their use is still a serious concern because military commanders rely on algorithmically generated recommendations for decision-making. Therefore, it is important that AI-DSS are employed lawfully and responsibly to ensure that decisions are not made hastily with devastating consequences.

While facial recognition offers a unique advantage by promoting accuracy in identifying or verifying an individual's identity, its effective use depends upon the environment this technology is employed in. FRT, when used in an uncontrolled environment, increases inaccuracy and the likelihood of false positives when identifying individuals in armed conflicts. Yet, it has not been determined by IHL how, when and to what extent FRT can be used to identify or verify individuals for targeting-purposes. The interplay between FRT and AI-DSS requires further clarification regarding to what extent they can be used to support military decision makers in targeting operations.

The use of AI-DSS raises unique legal concerns because of the scale and speed at which algorithmically-generated recommendations are generated. Given the speed at which AI operates, it raises concerns about how the fast pace of AI-DSS recommendations impact the human judgement of military commanders who rely on these recommendations to ensure compliance with IHL. As illustrated in the case study of the Lavender, AI-DSS can be used to perform certain functions, such as identification or labelling of potential targets which are delegated to these systems, which raises questions surrounding human judgement and responsibility.

Ensuring the lawful use of AI-DSS in armed conflicts requires thorough verification to confirm that recommended targets are both accurate and not protected from direct attack under IHL. There is a risk of over-emphasising the need for speedy decision-making at the cost of harm to the civilian population due to inaccuracy. To mitigate this risk and to comply with IHL obligations,

210 Ekelhof (n 1).

211 ICRC, 'International Humanitarian Law and the Challenges of Contemporary Armed Conflicts' (Geneva, November 2019).

it may be necessary to limit the role of AI-DSS to certain tasks related to the use of force, restrict its use in contexts with a high civilian presence, and slow down the military decision-making process. This will provide decision-makers with sufficient time to conduct qualitative assessments required by IHL obligations in targeting situations. Finally, the reporting of the Lavender illustrates the lack of transparency on this issue, which affects the ability of scholars to understand how militaries use AI-DSS in armed conflicts. States' secrecy about their use of AI raises serious concerns about accessing evidence for potential investigations and maintaining responsibility for IHL violations, and obstructs oversight of their design, development, and use. I call upon governments to be transparent about their use, policies, and regulations of AI-DSS in armed conflicts.

Acknowledgments

Many thanks to the reviewers for their helpful and insightful comments on earlier drafts of this article. This article has benefitted from comments received at the Conference on the Law Applicable to the Use of Biometrics by Armed Forces (Tallinn, 7–8 May 2024). I am grateful to Marten Zwanenburg and Sebastian Cymutta for their invaluable comments and feedback on this article.