Human-in-the-loop Dilemmas
The Lavender System in Israel Defence Forces Operations

Rohit Kumar Sharma*

In the first week of April 2024, *+972 Magazine* and *Local Call* reported on an artificial intelligence (AI) based programme called 'Lavender'. The magazine, published by a collective of Israeli-Palestinian journalists, uncovered troubling details concerning the deployment of 'Lavender' by the Israeli military in the ongoing conflict. According to the claims made by the report, the Lavender system played a central role in the unprecedented bombing in Gaza as Israel Defence Forces (IDF) used it to generate targets.[1] Moreover, the system reportedly generated thousands of 'potential' targets and 'suspected' militants based on the information fed to the machine about 'characteristics of known Hamas and [Palestinian Islamic Jihad] PIJ operatives'.[2]

Following the publication of the report, the IDF, in a swift response, denied using any AI system 'that identifies terrorist operatives or tries to predict whether a person is a terrorist'.[3] The IDF did acknowledge the existence of the 'database' to cross-reference intelligence sources that produce the latest information on the adversaries.[4] Amidst the allegations and denial, it is challenging to confirm the integrity of the report. However, given the unprecedented humanitarian crisis unfolding in Gaza, such allegations, even if unfounded cannot be dismissed or ignored.

* Mr Rohit Kumar Sharma is a Research Analyst at the Manohar Parrikar Institute for Defence Studies & Analyses (MP-IDSA), New Delhi.

ISSN 0976-1004 (print); 2583-7567 (online)

Journal of Defence Studies, Vol. 18, No. 2, April–June 2024, pp. 166–175

Such developments, in fact, offer an opportune moment to critically examine the promise of 'efficacy' and 'reliability' that entails the usage of AI systems on the battlefield. Moreover, the extent of destruction to infrastructure and human lives in the ongoing conflict undermines the argument that humans supervising these AI systems (weapon systems, decision-making, or decision support systems) can produce precise results with 'human-machine' synergy. Some observers have argued that these programmes are accelerating the 'kill chain' by *making the process of killing progressively more autonomous*.[5]

Against the backdrop of this report, the commentary focuses on the role of humans in operating AI systems on the battlefield and critically examines the challenges associated with 'human-machine' synergy. Given that humans play a crucial role at every stage of powering AI systems—from data collection and cleaning to making decisions based on system assessments—the commentary argues that keeping humans involved in AI systems presents its own set of challenges.

'Human–Machine Teaming'

In 2021, a book on AI was published under the pen name Brigadier General Y.S. The book titled *The Human Machine Team* begins with a primer on Machine Learning (ML) and its ability to learn and draw conclusions from big data to make predictions.[6] The book underscores the significance of human–machine synergy in attaining 'super cognition'. Elaborating further on the potential of the synergy, the later sections of the book explain how this arrangement is instrumental in creating a 'target in context' even before any battle begins. For decades, Brigadier General Y.S. notes, the armed forces of the world have been trying to connect the 'intelligence factory' with the 'fire factory'; the former represents an ability to leverage data to create more targets, while the latter denotes the ability to engage the targets at the right time and with suitable weapons.[7]

A recent report by *The Guardian* indicates that the author of 'Human Machine Team' is the commander of Unit 8200, an Israeli Intelligence Corps unit responsible for collecting signal intelligence (SIGINT). The radical vision shared by the commander in the book is, in fact, not an abstract notion. The IDF has been pioneering AI systems in its operation in Gaza, emulating what is prescribed in the book.[8] The real-world manifestation of the human–machine synergy is the Lavender system, currently in the public realm, for its questionable use by the IDF.

Operational Use of Lavender

Despite media reports on the use of the Lavender system, it is challenging to find a definite account expounding on its functioning. According to IDF's account, Lavender is a general-purpose database that organises and cross-references layers of existing intelligence sources.[9] The tool also helps in connecting different data points and complements existing processes for identifying adversaries. It can be qualified as an AI-decision support system that determines how information is 'processed, filtered, and presented' to end users.[10] The ML algorithm processes historical data to identify and generate targets based on the attributes and features of Hamas-linked operatives that were used to train the system.

The Lavender system is known to have identified as many as 37,000 Palestinians as suspected militants and their homes for possible targets.[11] As reported, the Lavender system is not an epitome of human–machine synergy that begets precision. When examining the system itself, noticeable gaps were apparent that the IDF failed to address before deploying it in the field. Firstly, the system was operationalised even when the error was approximately 10 per cent (also meaning 90 per cent accuracy) of the cases.[12] Secondly, in multiple instances, it was reported that the potential targets were marked despite them having a loose or no connection with the militant groups operating in the Gaza Strip. The practice partially explains how the system generated thousands of potential targets before the IDF commenced its ground offensive against Hamas.

Another AI system that was in the news for contentious reasons is 'The Gospel', which is primarily put to use to mark buildings and structures.[13] In the initial phase of the war, the IDF categorised the potential targets into tactical targets, underground targets, power targets and family/operative homes. The IDF believed that the categorisation would provide ease in prioritising the targets. Another contentious tool in use is the 'Where's Daddy' system, known to track targets and signal the operator when they enter their family homes.[14] The system also enabled targeting of the non-combatants that happened to be around the 'legitimate' targets.

Amidst all these developments and public discussions around these systems, one must refrain from associating these software programs solely with offensive operations, as the final decision to hit the target still lies with the operator. In fact, some scholars are of the opinion that systems like the Gospel are neither a weapon nor a decision-making system, and that it is more appropriate to categorise them as decision-support tools for commanders.[15]

Hence, despite the availability of these tools, the final decision to engage a target or to avoid it rests on the commanders.

Promise of Human-in-the-loop?

Sharing a historical anecdote in his book *Army of None,* Paul Scharre notes how Stanislav Petrov, a Soviet officer, saved the world from potential Armageddon as he correctly identified a system malfunction instead of perceiving it as incoming missiles from adversaries.[16] 'Humanity was saved', he writes, by a human-in-the-loop (HITL). Some scholars argue that despite discussions about Fully Autonomous Weapon Systems (FAWS) in future warfare, the increasing use of AI would increase the importance of humans in the war.[17] However, as reported in the ongoing conflict between Israel and Hamas, the promise of HITL does not seem to be working as efficient as expected.

With all discussions on the future of warfare focusing on the possibility of machines replacing humans, the current state of technology is far from achieving this. In fact, humans play a crucial role at every stage, both before and during the operationalisation of AI systems on the battlefield. At every stage of 'AI Data Hierarchy', humans remain the vital element. From data collection and storage by data engineers to data cleaning, testing, experimentation and deployment of AI models, humans play a critical role in decision-making.[18] Hence, it is pertinent to take into consideration the limitations that humans bring with them at each stage. These limitations come in the form of 'biases', 'emotions', 'moral deskilling', and also the uncertainty on the battlefield that mires the decision-making. Moreover, non-human/technological challenges are always present when designing and deploying such systems, especially on the battlefield. Therefore, both human and non-human limitations affect the assessments generated by AI systems, impacting decision-making in the battlefield.

Given the various sources of raw data generation, including Human Intelligence (HUMINT), SIGINT and Open Source Intelligence (OSINT), controlling and managing real-time data is challenging. At this stage, raw data is abundant. To make sense of enormous data, it is stored, which then undergoes a data cleaning process, a crucial step to ensure that data is accurate, consistent and free of errors.[19] Data cleaning ensures effective modelling and pattern recognition as the algorithm performs optimally when fed with high-quality data.

'Uncontrolled' conflict environments pose a wider range of challenges to the collection of complete and high-quality data.[20] Moreover, real-time

data collection during a conflict is not a simple process. Quality of data also varies by task and situation, and thereby, the 'microstructure' of the strategic environment is very important.[21] Therefore, data quality is strikingly different and more reliable in the case of administration and logistics than in combat reporting. Factors such as harsh conditions (dust, smoke, degraded sensors due to natural wear and tear, concealment of objects by adversaries, etc.), adversarial actions like spoofing attacks, and other battlefield surprises tend to degrade the quality of data, making it less effective for AI models. This fundamentally effects the AI systems, which consequentially effects decision-making in the military.

Reportedly, in the case of Lavender, information on characteristics of known Hamas and PIJ operatives was fed to the machine as training data, which was later used to locate these 'features' among the general population.[22] These characteristics are selected by humans (supervised classification), which later machine learns to identify on its own. It is also reported that the term 'Hamas' was loosely used to include even civil defence workers to train the dataset. Clearly, the data collected, stored and cleaned to train the machine had issues, as little to no consideration was given to distinguish the operatives of the militant group from non-combatants like the civil defence workers. This practice of collecting diverse data regarding the battlefield was alluded in the Human–Machine team, where the author stressed the significance of gleaning information about 'population, visual information, cellular data, social media connections, pictures, cellphone contacts, etc.'[23]. The effect of the type of data collected is also felt in the later stages of training, tuning and evaluation of the AI model.

The evaluation step is critical as it means bringing the AI model to real-world operations, meaning testing the model with data that has never been used for training. The operational experience helps the model 'improve', and the 'hyperparameters' are chosen and given more weight in the model.[24] Despite the multiple stages that a model goes through, the collected data remains the foundational pillar. The large number of reported casualties can therefore be attributed to the loose criteria used by the IDF to categorise a militant during data collection and model training.

It is also important to note that, unlike in the past, military personnel relying on AI to generate potential targets are unaware of how the machine reaches its decision. To operators, and even for the programmers, the AI model is the 'black box', which is incapable of providing a clear and understandable explanation of its decision-making process. When selecting a potential target, the inability to explain the logic of 'marking' a target can

give rise to unwanted casualties on the battlefield. However, the fact remains that the incapability stems from 'a general and inherent bug of the systems when it comes to AI'.[25]

The next stage, the most discussed phase, involves making final decisions on the potential target list generated by AI programs such as Lavender and Gospel systems. This is where the HITL aspect is most evident and frequently discussed. There is often a sense of comfort, as humans, not machines, are responsible for taking actions based on the AI model's output. However, the situation and the fact remains as complex as the AI systems itself. This phase of decision-making has its own challenges and limitations, which are worth pondering over.

Scholars have argued that the 'Prediction Machines' or the AI will have their most immediate impact at the decision level.[26] Explaining through an 'anatomy of a decision', Agrawal, Gans and Goldfarb argue that while prediction informs decision-making, other elements, such as judgement, data and action, remain, for now, firmly in the realm of humans.[27] Judgement is what determines the further course of action based on the prediction or output produced by AI. Elaborating further, Agrawal et al. argue that 'judgement is the process of determining the reward to a particular action in a particular environment…judgement [also] involves determining…the "reward function"'.[28] As with any decision that may have desirable or undesirable outcomes, the reward function entails relative rewards and penalties for each action that leads to specific results.[29]

Judgement also determines what is meaningful and what is at stake on a battlefield.[30] However, judgement in the military realm differs from that in the more stable civilian space. The military judgement involves considering the strategic environment, strategic missions, commanders' intent, rules of engagement, combat ethics and more.[31] Decision-making in the military, especially during a conflict, is neither straightforward nor easy. Despite this, since human lives are at stake, it is essential to consider actions carefully in light of the reward functions.

Reports indicate that the IDF's decision to attack targets, based on extensive lists generated by the machine, lacked due diligence and appropriate supervision due to over-reliance on the AI system. In some instances, the only human oversight was to confirm that the AI-selected target was male rather than female. These systems influenced the IDF to the extent that they treated the machine's output as if it were a human decision.

On this, Elke Schwarz notes that the cognitive load of extensive data and the cognitive disparity between a human brain and an AI system

leads to 'automation bias', which refers to humans' tendency to favour suggestions from automated decision-making systems over other sources.[32] 'When AI and human reasoning form an ecosystem, the possibility for human control is limited', notes Schwarz, sharing the grim assessment of human–machine teaming.[33] Elaborating on other pitfalls, she notes that these systems can also impose measurable costs to human performance, such as loss of situational awareness, skill degradation along with amplifying the 'responsibility gap'.[34]

Technology also adversely impacts the moral and ethical cultivation of humans during a conflict, as the mandate of speed and scale often conflicts with moral deliberation and considerations of actions. Against this backdrop, the humans in control of AI systems are more likely to privilege action over non-action in a time-sensitive conflict.[35] The AI–human configuration also leads to 'unjust erosion of targeting standards and morally devalues those subjected to violence'.[36] Discussing the technological risks of moral deskilling, Vallor states that military technologies provide the 'medium that defines what *kind* of soldier a soldier can choose to become, and what moral virtues she can develop through her service'.[37] The report shares some glaring examples where questionable decisions were made to engage a potential target. For instance, 'dumb' bombs instead of 'smart' precision bombs were used to target junior-level militants marked by Lavender.

Furthermore, the planners were less concerned about collateral damage. Reports indicate that at one point, killing 15 to 20 civilians was permissible for every junior-level militant. On several occasions, the army authorised the killing of more than 100 civilians to target senior-level Hamas officials with the rank of battalion or brigade commander. Clearly, the humans entrusted with the virtue of judging the reward function of their actions undermined the fundamental principles of International Humanitarian Law (IHL), namely, distinction and proportionality.[38] The principle of distinction requires that parties to the conflict must at all times distinguish between the civilian population/civilian objects and combatants. Similarly, the principle of proportionality prohibits military actions that could cause civilian casualties, injuries, or damage to civilian property when such harm would be excessive compared to the expected military gain.[39]

Some Israeli academics claim that the use of the tools in question is a 'very preliminary one in the chain of creating and authorising a military target'.[40] Sharing nuances about IDF's decision-making, Mimran and Dahan noted how an assessment made by an intelligence officer is delivered to a target room in which 'legal officers, operational advisors, engineers,

and more superior intelligence officers' take the final decision to engage or reject the target.[41] Despite all this, the IDF's conduct demonstrates beyond any reasonable doubt that the first wave of targeting decisions was done in great haste.[42]

Conclusion

Both supporters and opponents of deploying AI systems on the battlefield have their reasons. Supporters find comfort in retaining humans in the decision-making loop, seeing it as a necessary safeguard. Opponents, however, argue that deploying such systems has significant moral and ethical costs that may exacerbate the potential for dispassionate killings. The commentary examined the role of humans in light of reports on the Lavender system deployed by the IDF in the ongoing conflict in Gaza. Based on the anecdotes shared in the report, algorithm-informed suppositions prompted the target engagement without much deliberation at all. In fact, the decision-makers, on multiple occasions, failed to appreciate the existing IHL principles and acted in haste to respond to the machine-generated target lists. The military decisions made on the basis of Lavender's assessment were consequential in terms of human casualties. Despite human involvement in decisions to engage a target, the urgency with which potential targets were identified and subsequently acted upon is concerning. It raises questions about the role of humans in ensuring the safe and secure use of AI in the military.

Notes

1. Yuval Abraham, '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza', *+972 Magazine*, 3 April 2024, available at https://www.972mag.com/lavender-ai-israeli-army-gaza/.
2. Ibid.
3. 'Israel Defence Forces' Response to Claims about Use of 'Lavender' AI Database in Gaza', *The Guradian*, 3 April 2024, available at https://www.theguardian.com/world/2024/apr/03/israel-defence-forces-response-to-claims-about-use-of-lavender-ai-database-in-gaza#:~:text=IDF%20statement%20in%20response%20to,IDF%20directives%20and%20international%20law.
4. Ibid.
5. Elke Schwartz, 'Gaza War: Israel Using AI to Identify Human Targets Raising Fears That Innocents Are Being Caught in the Net', *The Conversation*, 12 April 2024, available at https://theconversation.com/gaza-war-israel-using-ai-to-identify-human-targets-raising-fears-that-innocents-are-being-caught-in-the-net-227422.

6. Brigadier General Y.S., *The Human Machine Team: How to Create Synergy Between Human & Artificial Intelligence That Will Revolutionize our World,* eBookPro Publishing, 2021 (Kindle edition).
7. Ibid., p. 51.
8. Harry Davies and Bethan McKernan, 'Op Israeli Spy Chief Exposes his True Identity in Online Security Lapse', *The Guardian,* 5 April 2024, available at https://www.theguardian.com/world/2024/apr/05/top-israeli-spy-chief-exposes-his-true-identity-in-online-security-lapse.
9. 'The IDF's Use of Data Technologies in Intelligence Processing', IDF, 18 June 2024, available at https://www.idf.il/en/mini-sites/idf-press-releases-regarding-the-hamas-israel-war/june-24-pr/the-idf-s-use-of-data-technologies-in-intelligence-processing/.
10. Klaudia Klonowska, 'Israel-Hamas 2024 Symposium – AI-based Targeting in Gaza: Surveying Expert Responses and Refining the Debate', Lieber Institute West Point, 7 June 2024, available at https://lieber.westpoint.edu/ai-based-targeting-gaza-surveying-expert-responses-refining-debate/.
11. '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza', no. 1.
12. Ibid.
13. Yuval Abraham, '"A Mass Assasination Factory": Inside Israel's Calculated Bombing of Gaza', *+972 Magazine*, 30 November 2023, available at https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/.
14. '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza', no. 1.
15. Tal Mimran, Magda Pacholska, Gal Dahan, and Lena Trabucco, 'Israel – Hamas 2024 Symposium – Beyond the Headlines: Combat Deployment of Military AI-based Systems by the IDF', Lieber Institute, 2 February 2024, available at https://lieber.westpoint.edu/beyond-headlines-combat-deployment-military-ai-based-systems-idf/.
16. Paul Scharre, *Army of None: Autonomous Weapons and the Future of War,* W.W.. Norton & Company, New York, 2018, p. 2.
17. Avi Goldfarb and Jon R. Lindsay, 'Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War', *International Security,* Vol. 46, No. 3.
18. Husanjot Chahal, Ryan Fedasiuk and Carrick Flynn, 'Messier than Oil: Assessing Data Advantage in Military AI', Center for Security and Emerging Technology (CSET), July 2020.
19. 'Overview of Data Cleaning', Geeks for Geeks, 24 May 2024, available at https://www.geeksforgeeks.org/data-cleansing-introduction/.
20. Arthur Holland Michel, 'Known Unknowns: Data Issues and Military Autonomous Systems', United Nations Institute for Disarmament Research (UNIDIR).
21. Avi Goldfarb and Jon R. Lindsay, 'Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War', no. 17, p. 21.
22. Yuval Abraham, '"Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza', no. 1.

23. Brigadier General Y.S, *The Human Machine Team: How to Create Synergy Between Human & Artificial Intelligence That Will Revolutionize our World*, no. 6, p. 71.
24. Ibid., p. 73.
25. Tal Mimran and Gal Dahan, 'Artificial Intelligence in the Battlefield: A Perspective from Israel', *OpinioJuris*, 20 April 2024, available at https://opiniojuris.org/2024/04/20/artificial-intelligence-in-the-battlefield-a-perspective-from-israel/.
26. Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence,* Harvard Business Review Press, Boston, 2022, p. 86.
27. Ibid., p. 88.
28. Ibid., p. 91.
29. Ibid.
30. Avi Goldfarb and Jon R. Lindsay, 'Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War', no. 17, p. 9.
31. Ibid., p. 22.
32. Sigal Samuel, 'Some Say AI Will Make War More Humane. Israel's War in Gaza Shows the Opposite', *Vox,* 8 May 2024, available at https://www.vox.com/future-perfect/24151437/ai-israel-gaza-war-hamas-artificial-intelligence.
33. Elke Schwarz, 'Silicon Valley Goes to War: Artificial Intelligence, Weapon Systems and the De-Skilled Moral Agent', *Philosophy Today,* Vol. 65, No. 3, 2021, pp. 549–569.
34. Ibid.
35. Neil C. Renic and Elke Schwarz, 'Inhuman-in-the-loop: AI-targeting and the Erosion of Moral Restraint', *OpinioJuris*, 19 December 2023, available at https://opiniojuris.org/2023/12/19/inhuman-in-the-loop-ai-targeting-and-the-erosion-of-moral-restraint/.
36. Ibid.
37. Shannon Vallor, 'Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character', *Philos. Technol.,* Vol. 28, 2015, pp. 107–124.
38. 'Fundamental Principles of IHL', International Committee of the Red Cross (ICRC), available at https://casebook.icrc.org/a_to_z/glossary/fundamental-principles-ihl.
39. 'Proportionality', ICRC, available at https://casebook.icrc.org/a_to_z/glossary/proportionality.
40. Tal Mimran and Gal Dahan, 'Artificial Intelligence in the Battlefield: A Perspective from Israel', no. 25.
41. Ibid.
42. Christopher Elliott, 'Expedient or Reckless? Reconciling Opposing Accounts of the IDF's Use of AI in Gaza', *OpinioJuris*, 26 April 2024, available at https://opiniojuris.org/2024/04/26/expedient-or-reckless-reconciling-opposing-accounts-of-the-idfs-use-of-ai-in-gaza/.